

系统辨识(8): 耦合辨识概念与方法

丁锋^{1,2,3}

摘要

耦合辨识是系统辨识的一个重要分支,是新近发展和提炼形成的一种辨识概念,主要用于研究结构复杂的参数耦合线性及非线性多变量系统的辨识问题. 辅助模型辨识思想、多新息辨识理论、递阶辨识原理、耦合辨识概念是本文作者提出的一些新的辨识研究思路、理念和方法,分别能够用于研究存在未知过程变量的不可测系统的辨识,能够提高辨识方法的收敛速度和参数估计精度,能够解决结构复杂、大规模多变量系统及参数耦合多变量系统的辨识问题、减小辨识算法的计算量.

首先介绍多变量系统耦合辨识概念,在此基础上讨论多变量系统的几种(全)耦合最小二乘辨识方法、(全)耦合随机梯度辨识方法、部分耦合随机梯度辨识方法、部分耦合最小二乘辨识方法等,最后说明耦合辨识方法可推广用于有色噪声干扰多变量系统的辨识,并列出了一些多变量系统模型结构,阐述了耦合辨识概念可以结合辅助模型辨识思想、多新息辨识理论、递阶辨识原理、迭代搜索原理(梯度迭代、最小二乘迭代、牛顿迭代)等来研究线性或非线性多变量系统的辨识问题.

关键词

迭代辨识; 递推辨识; 参数估计; FIR模型; 方程误差模型; CAR模型; CARMA模型; CARAR模型; CARARMA模型; 输出误差模型; OEMA模型; OEAR模型; 辅助模型辨识; 多新息辨识; 递阶辨识; 耦合辨识

中图分类号 TP273

文献标志码 A

收稿日期 2012-06-04

资助项目 国家自然科学基金(60973043)

作者简介

丁锋,男,博士,教授,博士生导师,主要从事系统辨识、过程建模、自适应控制方面的研究. fding@jiangnan.edu.cn

1 江南大学 物联网工程学院,无锡 214122

2 江南大学 控制科学与 Engineering 研究中心,无锡, 214122

3 江南大学 教育部轻工过程先进控制重点实验室,无锡 214122

0 引言

科学的发展是多层次的、多方位的,大到宏观宇宙太空的探索,小到微观粒子的划分. 科学创新需要与众不同的思想,科学发现可能源于大胆的假说、推断、猜测,甚至是一般人认为不可能的假设. 例如,牛顿的万有引力定律; 笔者提出的辅助模型辨识思想、多新息辨识理论、递阶辨识原理,以及本文的耦合辨识概念等. 再如,在经典运动物理学中,对于不同惯性参照系,荷兰物理学家亨德里克·洛伦兹(1853—1928年)推导出不同惯性参照系之间物理量的转换关系,在数学上表现为几个简单的方程式(类似于高中数学中的坐标平移或坐标变换方程). 这些简单的坐标变换方程后被称为著名的洛伦兹变换(Lorentz transformation). 洛伦兹变换也成为狭义相对论(special relativity)中的基本方程组.

在一个相对的空间里,阿尔伯特·爱因斯坦(1879—1955年)作了一个大胆的“绝对的”假设:不同参照系中光速是不变的,即光速是绝对的,或光速不变原理假设,从而导致相对论的诞生. 众所周知,坐在火车上的人与火车相对速度为零,不可解释骑在光速上的人与光线的速度仍然是光速,即光速没有相对性,相对论的解释是光速太高(这不需要解释,因为本身就是一种假设,这个绝对的假设作为唯一绝对的参照系). 光速不变就把不同参照系中的洛伦兹方程链接起来,进行简单的求解,就得出时间、长度、质量等在不同参照系中都是变化的. 这改变了人们的时空观. 所以说“洛伦兹变换公式”是爱因斯坦狭义相对论的数学基础. 爱因斯坦对相对论的主要贡献就是基于洛伦兹变换,假设光速是不变的. 这也是为什么有的学者评论“爱因斯坦是否偷了洛伦兹的思想?”

科学发现需要敏锐的洞察力、大胆的假说和假设,还需要坚忍的毅力. 20世纪70年代初,物理学家们普遍认为,世界上只有3种夸克,用3种夸克的理论就能够解释世界上所有的现象. 1974年,丁肇中教授提出了“寻找新粒子与新物质”的实验方案,可惜未能被多数物理学家们重视,但他执著地求索,最终在实验中发现了新粒子——J粒子,这一实验结果证明了当时3种夸克的理论是错误的. 这个发现,被国际高能物理学界誉为物理发展史上的一个重要里程碑. 丁肇中教授在回忆这段实验经历时说“做基础研究要有信心,做你认为正确的事,就要坚持去做,不要畏惧困难,不要因为多数人的反对而改

变,也不要管其他人怎么看,换句话说,要原谅大多数人的错误.科学是少数人推翻多数人的观念,科学是多数服从少数,只有少数人把多数人的观念推翻以后,科学才能向前发展,才能引领社会前进.”

记得20年前,笔者在清华大学攻读博士学位时,选择了“时变系统辨识”作为博士学位论文研究课题,得到笔者导师方崇智教授等老一辈过程辨识专家的反对,他们以及领域同行都认为系统辨识发展得很成熟,要想作出成绩极其困难,建议选择一个容易取得成果的课题.但笔者坚持了自己的选择,完全是出于对系统辨识的爱.这种爱使笔者能不断克服工作中的重重困难,并有坚强的毅力去面对和攻克一个个科学难题,取得一个个新的突破.尽管这种毅力使笔者失去早在1997年评上清华大学教授的机会,但令人愉悦的是对科学的贡献,得到的是对系统辨识的贡献—辅助模型辨识思想的创立、多新息辨识理论的提出、递阶辨识原理的诞生、耦合辨识概念的形成,以及用于随机时变系统参数估计算法和自适应控制算法收敛性研究的鞅超收敛定理的问世^[1-3].

笔者认为:科学发现、科学创新的方式是多样的,探索未知是一种创新,从小到大是一种创新(如发现遥远的巨大天体),从大到小也是一种创新(如发现构成物质的基本粒子、微粒子、微微粒子等).对研究建立系统数学模型的系统辨识学科来说,对于存在未知过程变量的系统来说,用系统的可测信息建立一个辅助模型,用辅助模型的输出代替系统的不可测未知变量,从而解决存在未知内部变量的系统辨识问题,这种新的研究思路就是辅助模型辨识思想.回顾标量代数到线性代数,再到矩阵代数、矩阵理论的发展历程,就是从小到大的过程.多新息辨识理论就是从小到大的例子:将系统的标量辨识新息扩展为新息向量,将向量新息扩展为新息矩阵,从而诞生了多新息辨识理论与多新息辨识方法.递阶辨识原理就是从大到小的例子:采用分解技术,将维数高、结构复杂的大规模系统的辨识问题化为维数低、相对容易的辨识问题进行求解,导致了递阶辨识方法,这就是递阶辨识原理产生的基础.耦合辨识概念就是基于一个合理假设,新的数据可以改进参数估计精度,将共有的子系统辨识算法中的参数估计连接起来,提出了耦合辨识方法,这就是耦合辨识概念的形成过程.

新思想、新理论、新原理、新概念的诞生都是科学史上的重要里程碑.就研究建立系统数学模型的

理论与方法的系统辨识而言,辅助模型辨识思想、多新息辨识理论、递阶辨识原理、耦合辨识概念的诞生,有助于推动系统辨识学科的研究进程^[4-11].笔者认为:新的方法,可以根据某种准则或原理推导出来,也可以采用某种近似,或类比,或某种合理的假设直接给出.这些新方法可能目前无法用严密的数学语言进行推导,倘若能证明这些方法的有效性,或仿真实验结果可行,或实际应用成功,那么不管在什么假设下、不管用什么途径提出的方法都是有效的^[12].

本文介绍一类新型的辨识方法:耦合辨识方法.它是基于本文作者提出的耦合辨识概念而建立的一些辨识方法.耦合辨识的第1篇重要研究论文发表在国际著名期刊《IEEE Transactions on Automatic Control》^[13]上.基于耦合辨识概念的耦合辨识方法可望发展为一个崭新的辨识领域,耦合辨识概念可以用于研究各种结构复杂的线性或非线性多变量系统的辨识问题.本文应用耦合辨识概念,讨论一些原创性耦合辨识方法,在本刊首次公开发表.本文较长,为便于阅读,特将本文框架结构列示如下.

- 0 引言
- 1 耦合辨识概念
- 2 全耦合辨识方法(C-LS)
 - 2.1 子系统最小二乘估计算法(SLS)
 - 2.2 耦合最小二乘估计算法(C-LS)
 - 2.3 耦合随机梯度估计算法(C-SG)
- 3 部分耦合随机梯度辨识方法(PC-SG)
 - 3.1 随机梯度辨识算法(SG)
 - 3.2 子系统随机梯度辨识算法(SSG)
 - 3.3 部分耦合随机梯度辨识算法(PC-SG)
- 4 部分耦合最小二乘辨识方法(PC-LS)
 - 4.1 最小二乘辨识算法(LS)
 - 4.2 子系统最小二乘辨识算法(SLS)
 - 4.3 部分耦合子系统最小二乘辨识算法(PC-SLS)
 - 4.4 部分耦合最小二乘辨识算法(PC-LS)
- 5 有色噪声干扰的多变量系统耦合辨识方法
 - 5.1 多变量伪线性回归系统的全耦合辨识方法
 - 5.2 类多变量方程误差类系统的部分耦合辨识方法
 - 5.3 类多变量输出误差类系统的部分耦合辨识方法

5.4 多变量方程误差类系统的耦合辨识方法

5.5 多变量输出误差类系统的耦合辨识方法

6 结语

1 耦合辨识概念

耦合辨识概念(coupled identification concept)主要用于研究结构复杂、子系统间存在参数耦合的线性或非线性的多变量系统辨识问题. 简单地说, 耦合辨识概念是研究参数耦合多变量系统辨识的理论与方法. 这里的耦合是指多变量系统的一些子系统辨识模型间的参数耦合关系, 说明多变量子系统辨识算法参数估计间的耦合关系, 并非指多变量系统不同通道输入和输出变量间的耦合关系.

对于标量线性系统, 或标量线性参数系统(一类特殊的非线性系统)^[5], 通常可以把系统的参数分离出来, 构成一个参数向量 $\theta \in \mathbf{R}^n$, 把系统的输入 $u(t)$ 、输出 $y(t)$ 等变量构成一个信息向量 $\varphi(t) \in \mathbf{R}^n$, 系统方程就可等价转换为下列单一的辨识模型:

$$y(t) = \varphi^T(t) \theta + v(t), \quad (1)$$

或

$$y(t) = \theta^T \varphi(t) + v(t).$$

其中 $v(t)$ 为零均值随机干扰噪声.

对于标量系统, 辨识模型只有一个, 不存在子系统间参数耦合关系. 对于多变量系统(即多输入多输出系统), 系统的输出有多个, 可以按照输出的数目分解为多个子系统, 每个子系统是一个多输入单输出系统. 假设有 m 个输出 $y_i(t)$ $i=1, 2, \dots, m$, 有 r 个输入 $u_j(t)$ $j=1, 2, \dots, r$, 就输出 $y_i(t)$ 构成的子系统而言, 同样可把子系统 i 的参数收集起来构成一个参数向量 $\theta_i \in \mathbf{R}^{n_i}$, 把系统的 r 个输入 $u_j(t)$ 和输出 $y_i(t)$ 等变量构成子系统 i 的信息向量 $\varphi_i(t) \in \mathbf{R}^{n_i}$, 子系统 i 也可等价转换为下列形式的一个辨识模型:

$$y_i(t) = \varphi_i^T(t) \theta_i + v_i(t). \quad (2)$$

其中 $v_i(t)$ 为第 i 个子系统的干扰噪声.

当 $i=1, 2, \dots, m$ 时, 从式(2)就得到 m 个辨识模型. 这 m 个子系统辨识模型之间的关系可能有下列几种情况.

1) 对于 m 个辨识模型(2), 子系统 i 的参数向量 θ_i 与子系统 j 的参数向量 θ_j ($i \neq j$) 是独立的, 子系统 i 的信息向量 $\varphi_i(t)$ 与子系统 j 的信息向量 $\varphi_j(t)$ ($i \neq j$) 也是独立的, 那么所有标量系统的辨识方法都可应用于子系统(2)的辨识.

2) 对于 m 个辨识模型(2), 子系统 i 的参数向量 θ_i 与子系统 j 的参数向量 θ_j ($i \neq j$) 是独立的, 子系统 i 与子系统 j 拥有部分相同的信息向量, 记作为 $\phi(t)$ ($i \neq j$), 不同部分的信息向量记作为 $\phi_i(t)$ 和 $\phi_j(t)$, 这时子系统 i 和子系统 j 辨识模型有下列形式:

$$y_i(t) = [\phi^T(t) \ \phi_i^T(t)] \theta_i + v_i(t), \quad (3)$$

$$y_j(t) = [\phi^T(t) \ \phi_j^T(t)] \theta_j + v_j(t). \quad (4)$$

它们可以等价写为

$$y_i(t) = [\phi^T(t) \ \phi_i^T(t)] \theta_i + v_i(t), \quad i=1, 2, \dots, m. \quad (5)$$

对于这种子系统拥有部分相同的子信息向量 $\phi(t)$ 的多变量系统, 需要研究新的辨识方法(包括递推方法、迭代方法、牛顿方法等), 以减少子系统参数辨识方法的计算量.

3) 对于 m 个辨识模型(2), 子系统 i 与子系统 j 有部分相同的参数向量 ϑ ($i \neq j$), 这时子系统 i 和子系统 j 辨识模型有下列形式,

$$y_i(t) = \varphi_i^T(t) \begin{bmatrix} \vartheta \\ \vartheta_i \end{bmatrix} + v_i(t), \quad (6)$$

$$y_j(t) = \varphi_j^T(t) \begin{bmatrix} \vartheta \\ \vartheta_j \end{bmatrix} + v_j(t). \quad (7)$$

它们可以等价写为

$$y_i(t) = \varphi_i^T(t) \begin{bmatrix} \vartheta \\ \vartheta_i \end{bmatrix} + v_i(t), \quad i=1, 2, \dots, m. \quad (8)$$

这个模型包含了下列特殊情况:

$$y_i(t) = [\varphi_i^T(t) \ \varphi_i^T(t)] \begin{bmatrix} \vartheta \\ \vartheta_i \end{bmatrix} + v_i(t), \quad i=1, 2, \dots, m. \quad (9)$$

4) 对于 m 个辨识模型(2), 最特殊的情况是所有子系统拥有一个共同的参数向量 θ , 这时辨识模型(2)可以写为

$$y_i(t) = \varphi_i^T(t) \theta + v_i(t), \quad i=1, 2, \dots, m. \quad (10)$$

如果进一步令

$$y(t) := \begin{bmatrix} y_1(t) \\ y_2(t) \\ \vdots \\ y_m(t) \end{bmatrix} \in \mathbf{R}^m, \quad \Phi(t) := \begin{bmatrix} \varphi_1^T(t) \\ \varphi_2^T(t) \\ \vdots \\ \varphi_m^T(t) \end{bmatrix} \in \mathbf{R}^{m \times n},$$

$$v(t) := \begin{bmatrix} v_1(t) \\ v_2(t) \\ \vdots \\ v_m(t) \end{bmatrix} \in \mathbf{R}^m. \quad (11)$$

那么式(10)可以等价写为

$$y(t) = \Phi(t) \theta + v(t). \quad (12)$$

这是一个多变量线性回归模型(multivariable linear regression model)或多元线性回归模型(multi-

variate linear regression model)^[5]. 这个模型很有用途. 特别有意义. 周期时变系统可采用这个辨识模型. 批过程(batch process) 重复控制系统也可采用这个辨识模型. 因为对于同一个时不变参数系统(可以是标量系统), 使用不同的输入信号序列多次激励这个系统, 也就是说, 进行重复实验, 得到输入输出数据, 就可用这个多元辨识模型.

这个模型还可以用于迭代学习控制系统(iterative learning control system) 的辨识^[14]. 迭代学习控制是一种重复控制, 且第 $k+1$ 批次时刻 t 的控制信号 $u_{k+1}(t)$ 是在前一次 $u_k(t)$ 基础上用输出跟踪误差 $e_k(t+1)$ 加以修正, 控制的目标是设计一个输入信号, 使系统的输出跟踪期望的输出. 迭代重复学习控制一般采用离散时间模型(因为需要记录先前的输入信号), 如

$$A(z) y_k(t) = B(z) u_k(t) + v_k(t). \quad (13)$$

这里 $v_k(t)$ 为噪声, 系统延迟为 d , $\{u_k(t)\}$ 和 $\{y_k(t)\}$ 分别是系统的输入和输出序列, z^{-1} 为单位后移算子 [$z^{-1}y(t) = y(t-1)$, $zy(t) = y(t+1)$], $A(z)$ 和 $B(z)$ 是单位后移算子 z^{-1} 的常系数时不变多项式:

$$A(z) := 1 + a_1 z^{-1} + a_2 z^{-2} + \dots + a_{n_a} z^{-n_a},$$

$$B(z) := b_1 z^{-1} + b_2 z^{-2} + \dots + b_{n_b} z^{-n_b}.$$

假设 $A(z)$ 稳定, 否则先使用反馈使系统稳定, 然后将迭代学习控制用于闭环系统. 假设采样周期为 T , 考虑 L 个输入、输出样本:

$$u_k(t) \quad t = lT \quad l \in \{0, 1, \dots, L-1\},$$

$$y_k(t) \quad t = lT \quad l \in \{d, d+1, \dots, d+L-1\}.$$

期望输出为

$$y_d(t) \quad t = lT \quad l \in \{d, d+1, \dots, d+L-1\}.$$

系统的跟踪性能是以误差信号 $e_k(t) := y_d(t) - y_k(t)$ 的大小来评价的. 实际中, 系统的运行时间长度 L 总是有限的, 尽管有时为了分析和设计总是假设 L 是无穷大. 迭代次数 k 可以是无限的 $k \in \{0, 1, 2, \dots\}$, 但更希望学习很有限的次数, 系统的跟踪误差令人满意. d 是严格真传递函数 $G(z) := B(z)/A(z)$ 的分母分子相对阶次, 在这里假设相对阶次为 1, 即 $d=1$. 广泛使用的迭代学习算法为

$$u_{k+1}(t) = Q(z) [u_k(t) + R(z) e_k(t+1)],$$

其中 $Q(z)$ 和 $R(z)$ 是 z^{-1} 的常系数时不变多项式, 分别定义为 Q -滤波器和学习函数^[14]. 当 $k=1, 2, \dots, m$ 时, 参数化系统(13), 像式(11)那样定义相关向量和矩阵, 就可得到式(12)的辨识模型.

这里笔者要强调一下, 迭代学习控制中, 有一点在很多文献上没有论述清楚: 式(13)表达不清晰. 大家注意, 由于有 $k=1, 2, \dots$ 的存在, 式(13)中的 t 是不同参照系的时间. 比如时间以“d”为单位(即 24 h), 系统每天模仿同一个动作的控制, $k=1$ 代表今天(第 1 天), 那么 $t=0$ 代表今天凌晨零点开始; $k=2$ 代表明天(第 2 天), 那么对应的 $t=0$ 代表明天凌晨零点开始, 依次类推. 在这种情况下, 迭代学习控制系统模型(13)应该修正为

$$A(z) y(24k+t) = B(z) u(24k+t) + v(24k+t), \quad k=0, 1, 2, \dots \quad (14)$$

且 $t=lT$ 只能是有限的, 期望输出 $y_d(t)$ 中的 t 是每一个参照系的时间, 精确表述就应该修改为 $y_d(24k+t)$, 由于是每天模仿同一个动作, 因而是一个周期函数 $y_d(24+t) = y_d(t)$, 而 t 的取值只能是有限的 $0 \leq t < 24$. 这相当于期望输出 $y_d(t)$ 是一个周期函数. 因此, 迭代学习控制的期望输出可认为是一周期函数.

5) 对于 m 个辨识模型(2), 子系统 i 与子系统 j 有部分相同的参数向量, 记作 $\alpha (i \neq j)$, 不同部分的参数向量记为 θ_i , 子系统 i 与子系统 j 拥有部分相同的信息向量, 记作 $\varphi(t) (i \neq j)$, 不同部分的信息向量记为 $-\psi_i(t)$, 这时子系统 i 和子系统 j 辨识模型有下列形式:

$$y_i(t) = [-\psi_i(t) \quad \varphi^T(t)] \begin{bmatrix} \alpha \\ \theta_i \end{bmatrix} + v_i(t), \quad (15)$$

$$y_j(t) = [-\psi_j(t) \quad \varphi^T(t)] \begin{bmatrix} \alpha \\ \theta_j \end{bmatrix} + v_j(t), \quad (16)$$

或等价写为

$$y_i(t) = [-\psi_i(t) \quad \varphi^T(t)] \begin{bmatrix} \alpha \\ \theta_i \end{bmatrix} + v_i(t) =$$

$$-\psi_i(t) \alpha + \varphi^T(t) \theta_i + v_i(t) =$$

$$-\psi_i(t) \alpha + \theta_i^T \varphi(t) + v_i(t), \quad i=1, 2, \dots, m. \quad (17)$$

令

$$\Psi(t) := \begin{bmatrix} \psi_1(t) \\ \psi_2(t) \\ \vdots \\ \psi_m(t) \end{bmatrix} \in \mathbf{R}^{m \times n}, \quad \Theta := [\theta_1 \quad \theta_2 \quad \dots \quad \theta_m] \in \mathbf{R}^{n \times m}.$$

则式(17)可以等价写为

$$\begin{bmatrix} y_1(t) \\ y_2(t) \\ \vdots \\ y_m(t) \end{bmatrix} = \begin{bmatrix} -\psi_1(t) \alpha \\ -\psi_2(t) \alpha \\ \vdots \\ -\psi_m(t) \alpha \end{bmatrix} + \begin{bmatrix} \theta_1^T \varphi(t) \\ \theta_2^T \varphi(t) \\ \vdots \\ \theta_m^T \varphi(t) \end{bmatrix} + \begin{bmatrix} v_1(t) \\ v_2(t) \\ \vdots \\ v_m(t) \end{bmatrix},$$

即

$$y(t) = -\psi(t)\alpha + \theta^T \varphi(t) + v(t),$$

或

$$y(t) + \psi(t)\alpha = \theta^T \varphi(t) + v(t).$$

这就是拥有一个公分母作为特征多项式的多变量系统输入输出表达 (input-output representation) 的辨识模型或传递函数阵主模型的多变量系统辨识模型^[15-21] 其特征是系统包含了一个参数向量 α 和一个参数矩阵 θ . 后面还要讨论这类多变量系统的耦合辨识问题.

对于第 2 种情形, 即子系统部分信息向量相同的辨识模型 (5), 可以研究部分信息向量耦合的辨识方法; 对于后 3 种情形, 即子系统部分参数耦合辨识模型 (8), 子系统参数全耦合辨识模型 (10), 以及子系统部分参数耦合与部分信息向量耦合辨识模型 (17), 由于子系统间存在参数耦合关系, 独立的子系统辨识方法计算量大, 所以必须借助于耦合辨识概念, 研究新的耦合辨识方法, 以解决子系统辨识方法计算量大的问题.

耦合辨识概念是继笔者提出和建立辅助模型辨识思想、多新息辨识理论、递阶辨识原理后, 新近提炼形成的一种辨识理念和方法. 它可以用于参数耦合多变量系统的辨识, 可以结合现存的辨识方法 (如递推最小二乘辨识方法^[6]、辅助模型辨识方法^[7]、多新息辨识方法^[9]、递阶辨识方法 (两阶段或多阶段辨识方法)^[5]、迭代辨识方法^[8]、基于滤波的辨识方法^[22]、牛顿方法^[23] 等, 研究各种干扰噪声模型的多变量系统辨识问题, 提出相应的耦合辨识方法.

耦合辨识概念是笔者 1988 年在清华大学学习《过程辨识》^[24] 中多变量线性过程的参数估计方法时, 受到具有公分母特征值多项式传递函数阵的多变量系统递推最小二乘估计方法, 以及子系统辨识方法的启发 (后来建立了递阶辨识原理, 并提出了一系列递阶辨识方法^[15-16, 25-26]), 经过多年的思考提炼出的. 第 1 篇耦合辨识重要研究论文 “Partially coupled stochastic gradient identification methods for non-uniformly sampled systems” 见文献 [13].

文献 [24] 的方法实际上是本文的全耦合递推最小二乘辨识方法, 已经证明全耦合递推最小二乘辨识方法只是多变量系统递推最小二乘算法的不同表现形式, 只是避免矩阵求逆. 本文发展和研究了多变量全耦合随机梯度辨识方法、部分耦合最小二乘方法、部分耦合随机梯度辨识方法等一些原创性辨识

方法, 首次在本刊 “系统辨识” 连载最后一期发表. 后继将出版 “辨识方法的计算效率” 连载论文.

2 全耦合辨识方法 (C-LS)

首先定义一些符号. $\mathbf{1}_n$ 是元均为 1 的 n 维列向量; I 为适当维数的单位阵; I_n 为 n 阶单位阵; 范数定义为 $\|X\|^2 = \text{tr}[XX^T]$.

耦合辨识方法分为部分耦合辨识方法与全耦合辨识方法. 全耦合辨识方法简称为耦合辨识方法. 本节讨论全耦合辨识方法. 耦合辨识方法就是各子系统拥有一个共同的参数向量, 这个参数向量的估计在每一个子系统辨识方法间形成耦合关系. 这样的辨识方法称为耦合辨识方法. 如果子系统辨识方法间的所有参数估计都耦合起来, 就称为 (全) 耦合辨识方法 (coupled identification method); 如果子系统辨识方法间只有部分参数估计耦合起来, 就称为部分耦合辨识方法 (partially coupled identification method).

考虑下列多变量线性回归系统

$$y(t) = \Phi(t)\theta + v(t), \quad (19)$$

其中 $y(t) = [y_1(t) \ y_2(t) \ \dots \ y_m(t)]^T \in \mathbf{R}^m$ 为 m 维系统输出向量, $\Phi(t) \in \mathbf{R}^{m \times n}$ 是由系统输入输出数据构成的回归信息矩阵, $\theta \in \mathbf{R}^n$ 是待辨识的系统参数向量, $v(t) = [v_1(t) \ v_2(t) \ \dots \ v_m(t)]^T \in \mathbf{R}^m$ 是零均值白噪声向量. 假设 $t \leq 0$ 时 $y(t) = 0$, $\Phi(t) = 0$ 和 $v(t) = 0$.

令 $\hat{\theta}(t)$ 为参数向量 θ 在时刻 t 的估计. 定义和极小化梯度准则函数

$$J(\theta) = \|y(j) - \Phi(j)\theta\|^2$$

或最小二乘准则函数

$$J(\theta) = \sum_{j=1}^t \|y(j) - \Phi(j)\theta\|^2,$$

可以得到下列递推辨识算法

$$\hat{\theta}(t) = \hat{\theta}(t-1) + L(t) [y(t) - \Phi(t)\hat{\theta}(t-1)],$$

其中 $L(t) \in \mathbf{R}^{n \times m}$ 为增益矩阵.

当取增益向量为

$$L(t) = \frac{\Phi^T(t)}{R(t)} R(t) = R(t-1) + \|\Phi(t)\|^2,$$

$$R(0) = 1,$$

就得到了估计参数向量 θ 的随机梯度 (SG) 算法,

$$\hat{\theta}(t) = \hat{\theta}(t-1) + L(t) [y(t) - \Phi(t)\hat{\theta}(t-1)],$$

$$\hat{\theta}(0) = \mathbf{1}_n / p_0, \quad (20)$$

$$L(t) = \frac{\Phi^T(t)}{R(t)}, \quad (21)$$

$$R(t) = R(t-1) + \|\Phi(t)\|^2, \quad R(0) = 1. \quad (22)$$

当取增益向量为 $L(t) := P(t) \Phi^T(t)$ 就得到了估计参数向量 θ 的递推最小二乘(RLS)算法

$$\hat{\theta}(t) = \hat{\theta}(t-1) + L(t) [y(t) - \Phi(t) \hat{\theta}(t-1)],$$

$$\hat{\theta}(0) = \mathbf{1}_n/p_0, \quad (23)$$

$$L(t) = P(t) \Phi^T(t), \quad (24)$$

$$P^{-1}(t) = P^{-1}(t-1) + \Phi^T(t) \Phi(t), \quad P(0) = p_0 I_n. \quad (25)$$

为避免求式(25)大矩阵 $P(t) \in \mathbf{R}^{n \times n}$ 的逆(一般 $n > m$) ,应用矩阵求逆引理

$$(A + BC)^{-1} = A^{-1} - A^{-1}B(I + CA^{-1}B)^{-1}CA^{-1} \quad (26)$$

于式(25) ,得到等价递推最小二乘算法

$$\hat{\theta}(t) = \hat{\theta}(t-1) + L(t) [y(t) - \Phi(t) \hat{\theta}(t-1)],$$

$$\hat{\theta}(0) = \mathbf{1}_n/p_0, \quad (27)$$

$$L(t) = P(t-1) \Phi^T(t) [I_m + \Phi(t) P(t-1) \Phi^T(t)]^{-1}, \quad (28)$$

$$P(t) = [I_n - L(t) \Phi(t)] P(t-1), \quad P(0) = p_0 I_n. \quad (29)$$

递推算法中 ,参数估计向量初值一般取为零或很小实向量 ,如 $\hat{\theta}(0) = \mathbf{1}_n/p_0$, $p_0 = 10^6$. 多变量 RLS 算法(27) — (29) 中各变量的维数如表 1 所示.

在每一步递推过程中 ,多变量 RLS 算法都需要计算矩阵逆 $[I_m + \Phi(t) P(t-1) \Phi^T(t)]^{-1} \in \mathbf{R}^{m \times m}$,使得计算复杂性增加. 特别对于大 m (m 为输出的维数) ,计算量就很大. 这是多变量 RLS 算法(27) — (29) 的缺点 ,也激发人们研究新的耦合参数估计算法.

表 1 多变量 RLS 算法中各变量维数

Table 1 The dimensions of the variables in the RLS algorithm

变量名称	维数
输出向量	$y(t) \in \mathbf{R}^m$
参数向量	$\theta \in \mathbf{R}^n$
参数估计向量	$\hat{\theta}(t) \in \mathbf{R}^n$
信息矩阵	$\Phi(t) \in \mathbf{R}^{m \times n}$
增益矩阵	$L(t) \in \mathbf{R}^{n \times m}$
协方差阵	$P(t) \in \mathbf{R}^{n \times n}$

2.1 子系统最小二乘估计算法(SLS)

下面基于最小二乘辨识原理推导耦合最小二乘辨识方法.

令 $\phi_i^T(t) \in \mathbf{R}^{1 \times n}$ 是 $\Phi(t)$ 的第 i 行 ,即

$$\Phi(t) := \begin{bmatrix} \phi_1^T(t) \\ \phi_2^T(t) \\ \vdots \\ \phi_m^T(t) \end{bmatrix} \in \mathbf{R}^{m \times n}.$$

式(19)可以分解为 m 个辨识模型(子系统):

$$y_i(t) = \phi_i^T(t) \theta + v_i(t) \quad i=1, 2, \dots, m. \quad (30)$$

它们的每一个都包含了一个共同的参数向量 $\theta \in \mathbf{R}^n$. 很明显 ,只需其中一个子系统就可以辨识出参数向量 θ . 然而 ,必须利用所有子系统(即所有采集到的观测数据)来辨识 θ 以提高参数估计精度.

根据最小二乘原理 ,从式(30)可以获得 m 个 RLS 算法 ,即子系统最小二乘(SLS)算法:

$$\hat{\theta}_i(t) = \hat{\theta}_i(t-1) + P_i(t) \phi_i(t) [y_i(t) - \phi_i^T(t) \hat{\theta}_i(t-1)],$$

$$\hat{\theta}_i(0) = \mathbf{1}_n/p_0, \quad (31)$$

$$P_i^{-1}(t) = P_i^{-1}(t-1) + \phi_i(t) \phi_i^T(t),$$

$$P_i(0) = p_0 I_n, \quad i=1, 2, \dots, m. \quad (32)$$

其中 $P_i(t) \in \mathbf{R}^{n \times n}$ 是第 i 个子系统的协方差阵.

注意到每个子系统最小二乘算法(31) — (32) 用了一个共同的参数估计向量 $\hat{\theta}(t)$,但是每个子系统的参数估计 $\hat{\theta}_i(t)$ 是独立的 ,即子系统 i 的参数估计与子系统 j 的参数估计 $\hat{\theta}_j(t)$ 无关 ($i \neq j$) . 为清晰起见 ,将(31)中第 i 个子系统的参数向量 $\hat{\theta}_i(t)$ 记作 $\hat{\theta}_i(t)$,所以第 i 子系统 SLS 算法(31) — (32) 可以等价写为

$$\hat{\theta}_i(t) = \hat{\theta}_i(t-1) + P_i(t) \phi_i(t) [y_i(t) - \phi_i^T(t) \hat{\theta}_i(t-1)], \quad \hat{\theta}_i(0) = \mathbf{1}_n/p_0, \quad (33)$$

$$P_i^{-1}(t) = P_i^{-1}(t-1) + \phi_i(t) \phi_i^T(t),$$

$$P_i(0) = p_0 I_n, \quad i=1, 2, \dots, m. \quad (34)$$

应用矩阵求逆引理 ,可以得到等价的子系统 SLS 算法:

$$\hat{\theta}_i(t) = \hat{\theta}_i(t-1) + L_i(t) [y_i(t) - \phi_i^T(t) \hat{\theta}_i(t-1)],$$

$$\hat{\theta}_i(0) = \mathbf{1}_n/p_0, \quad (35)$$

$$L_i(t) = P_i(t) \phi_i(t) = P_i(t-1) \phi_i(t) [1 + \phi_i^T(t) P_i(t-1) \phi_i(t)]^{-1}, \quad (36)$$

$$P_i(t) = [I_n - L_i(t) \phi_i^T(t)] P_i(t-1),$$

$$P_i(0) = p_0 I_n, \quad i=1, 2, \dots, m. \quad (37)$$

这意味着子系统参数估计算法 $\hat{\theta}_i(t)$ 间没有任何耦合 ,参见 SLS 算法(33) — (34) 的方块图(图 1).

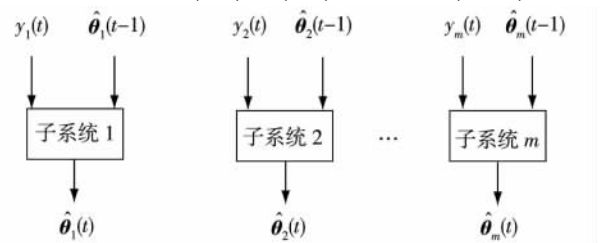


图 1 子系统最小二乘算法示意

Fig. 1 The schematic diagram of the subsystem least squares algorithm

注 1 对于 $i = 1, 2, \dots, m$, 从 (33) — (34) 获得 m 个参数估计向量 $\hat{\theta}_i(t)$, 它们都是共同参数向量 θ 的估计. 这导致参数向量 θ 的大量冗余估计, 因为每个 SLS 算法都对 θ 进行了估计. 一种办法是采用其平均作为 θ 的估计:

$$\hat{\theta}(t) = \frac{\hat{\theta}_1(t) + \hat{\theta}_2(t) + \dots + \hat{\theta}_m(t)}{m} \in \mathbf{R}^n. \quad (38)$$

如果把式 (38) 的参数估计 $\hat{\theta}(t)$ 仅仅作为参数向量 θ 的输出, 那么各子系统辨识算法仍是独立的, 其参数估计性能 ($\hat{\theta}_i(t)$ 的收敛性) 与多变量系统递推最小二乘法性能类似, 因为各子系统参数估计互不相关. 一般认为平均值估计 $\hat{\theta}(t)$ 是“最好”的估计, 所以可以用 $\hat{\theta}(t-1)$ 代替 SLS 算法 (33) — (34) 或 (35) — (37) 中的 $\hat{\theta}_i(t-1)$, 就得到一个简单的耦合子系统最小二乘 (C-SLS) 算法:

$$\hat{\theta}_i(t) = \hat{\theta}(t-1) + P_i(t) \phi_i(t) [y_i(t) - \phi_i^T(t) \hat{\theta}(t-1)], \quad \hat{\theta}(0) = \mathbf{1}_n/p_0, \quad (39)$$

$$P_i^{-1}(t) = P_i^{-1}(t-1) + \phi_i(t) \phi_i^T(t), \quad P_i(0) = p_0 \mathbf{I}_n, \quad i = 1, 2, \dots, m, \quad (40)$$

$$\hat{\theta}(t) = \frac{\hat{\theta}_1(t) + \hat{\theta}_2(t) + \dots + \hat{\theta}_m(t)}{m}. \quad (41)$$

或

$$\hat{\theta}_i(t) = \hat{\theta}(t-1) + L_i(t) [y_i(t) - \phi_i^T(t) \hat{\theta}(t-1)], \quad \hat{\theta}(0) = \mathbf{1}_n/p_0, \quad (42)$$

$$L_i(t) = P_i(t) \phi_i(t) = P_i(t-1) \phi_i(t) [1 + \phi_i^T(t) P_i(t-1) \phi_i(t)]^{-1}, \quad (43)$$

$$P_i(t) = [I_n - L_i(t) \phi_i^T(t)] P_i(t-1), \quad P_i(0) = p_0 \mathbf{I}_n, \quad i = 1, 2, \dots, m, \quad (44)$$

$$\hat{\theta}(t) = \frac{\hat{\theta}_1(t) + \hat{\theta}_2(t) + \dots + \hat{\theta}_m(t)}{m}. \quad (45)$$

这个耦合子系统递推最小二乘算法的收敛性有待进一步研究. 下面推导耦合最小二乘算法来避免冗余估计.

2.2 耦合最小二乘估计算法 (C-LS)

对于递推参数估计算法, 期望其是收敛的, 即假设参数估计随着数据长度的增大而收敛于真参数. 也就是说, 可以认为第 $i-1$ 个子系统在时刻 t 的参数估计 $\hat{\theta}_{i-1}(t)$ 比第 i 个子系统在时刻 $t-1$ 的参数估计 $\hat{\theta}_i(t-1)$ 更接近真参数 θ [10, 27-31].

参考文献 [13] 中的部分耦合随机梯度辨识方法, 并借助于 Jacobi 和 Gauss-Seidel 迭代思想 [32], 用 $\hat{\theta}_{i-1}(t)$ 代替式 (33) 右边的 $\hat{\theta}_i(t-1)$, $i = 2, 3, \dots, m$, 用 $\hat{\theta}_m(t-1)$ 代替式 (33) $i = 1$ 时右边的 $\hat{\theta}_1(t-1)$, 则

得到下列参数估计与协方差阵耦合最小二乘辨识算法 (Coupled Least Squares identification algorithm, C-LS) [24, 33]:

$$\hat{\theta}_i(t) = \hat{\theta}_{i-1}(t) + L_i(t) [y_i(t) - \phi_i^T(t) \hat{\theta}_{i-1}(t)] \quad (46)$$

$$L_i(t) = P_i(t) \phi_i(t), \quad (47)$$

$$P_i^{-1}(t) = P_i^{-1}(t-1) + \phi_i(t) \phi_i^T(t), \quad i = 2, 3, \dots, m, \quad (48)$$

$$\hat{\theta}_1(t) = \hat{\theta}_m(t-1) + L_1(t) [y_1(t) - \phi_1^T(t) \hat{\theta}_m(t-1)], \quad \hat{\theta}_m(0) = \mathbf{1}_n/p_0, \quad (49)$$

$$L_1(t) = P_1(t) \phi_1(t), \quad (50)$$

$$P_1^{-1}(t) = P_m^{-1}(t-1) + \phi_1(t) \phi_1^T(t), \quad P_m(0) = p_0 \mathbf{I}_n. \quad (51)$$

应用矩阵求逆引理 (26) 于式 (48) 和 (51), 则 C-LS 算法可以等价表达为

$$\hat{\theta}_i(t) = \hat{\theta}_{i-1}(t) + L_i(t) [y_i(t) - \phi_i^T(t) \hat{\theta}_{i-1}(t)] \quad (52)$$

$$L_i(t) = P_{i-1}(t) \phi_i(t) / [1 + \phi_i^T(t) P_{i-1}(t) \phi_i(t)], \quad (53)$$

$$P_i(t) = [I - L_i(t) \phi_i^T(t)] P_{i-1}(t), \quad i = 2, 3, \dots, m, \quad (54)$$

$$\hat{\theta}_1(t) = \hat{\theta}_m(t-1) + L_1(t) [y_1(t) - \phi_1^T(t) \hat{\theta}_m(t-1)], \quad \hat{\theta}_m(0) = \mathbf{1}_n/p_0, \quad (55)$$

$$L_1(t) = P_m(t-1) \phi_1(t) / [1 + \phi_1^T(t) P_m(t-1) \phi_1(t)], \quad (56)$$

$$P_1(t) = [I - L_1(t) \phi_1^T(t)] P_m(t-1), \quad P_m(0) = p_0 \mathbf{I}_n. \quad (57)$$

上述算法中, 子系统辨识算法间参数估计 $\hat{\theta}_i(t)$ 和协方差阵 $P_i(t)$ 都是耦合的, $\hat{\theta}_i(t) \in \mathbf{R}^n$, $L_i(t) \in \mathbf{R}^n$ 和 $P_i(t) \in \mathbf{R}^{n \times n}$ 分别为第 i 个子系统时刻 t 的参数估计向量、增益向量和协方差阵.

C-LS 算法 (52) — (57) (图 2) 计算参数估计向量 $\hat{\theta}_m(t)$ 的步骤如下.

1) 置初值. 令 $t = 1$, $\hat{\theta}_m(0) = \mathbf{1}_n/p_0$, $P_m(0) = p_0 \mathbf{I}_n$, $p_0 = 10^6$.

2) 收集观测数据 $y(t)$ 和 $\Phi(t)$, 设 $\phi_i^T(t) \in \mathbf{R}^{1 \times n}$ 为 $\Phi(t)$ 的第 i 行.

3) 用式 (56) 计算增益向量 $L_i(t)$, 用式 (57) 计算协方差阵 $P_i(t)$, 用式 (55) 刷新参数估计向量 $\hat{\theta}_1(t)$.

4) 对于每一个 $i (i = 2, 3, \dots, m)$, 用式 (53) 计算增益向量 $L_i(t)$, 用式 (54) 计算协方差阵 $P_i(t)$, 用式 (52) 刷新参数估计向量 $\hat{\theta}_i(t)$.

5) t 增 1, 转到第 2 步.

注 2 多变量 RLS 算法 (27) — (29) 要求计算矩阵逆 $[I_m + \Phi(t) P(t-1) \Phi^T(t)]^{-1}$, 但是 C-LS 算法

(52) — (57) 不涉及计算矩阵逆. 因此, C-LS 算法优于多变量 RLS 算法(27) — (29).

笔者已经证明: 耦合最小二乘辨识算法(52) — (57) 等价于多变量 RLS 算法(27) — (29), 且两个算法中的参数估计和协方差阵有关系: $\hat{\theta}(t) = \hat{\theta}_m(t)$, $P(t) = P_m(t)$ [24, 33]. 也就是说, 多变量 RLS 算法(27) — (29) 与耦合最小二乘辨识算法(52) — (57) 产生的参数估计是一样的, 只是形式上的差别.

正如笔者在引言中指出的: 科学创新需要与与众不同的思想. 科学发现可能源于大胆的、甚至是不可能的假设, 或一些改变. 在此, 对与多变量 RLS 算法等价的耦合最小二乘辨识算法进行改造, 提出新的耦合辨识方法. 注意到耦合最小二乘辨识算法(52) — (57) 中, 子系统辨识算法间参数估计 $\hat{\theta}_i(t)$ 和协方差阵 $P_i(t)$ 都是耦合的, 去掉子系统辨识算法(52) — (57) 间协方差阵的耦合, 便得到一个新的参数估计全耦合辨识算法:

$$\begin{aligned} \hat{\theta}_1(t) &= \hat{\theta}_m(t-1) + L_1(t) [y_1(t) - \phi_1^T(t) \hat{\theta}_m(t-1)], \\ \hat{\theta}_m(0) &= \mathbf{1}_n / p_0, \end{aligned} \quad (58)$$

$$\begin{aligned} L_1(t) &= P_1(t-1) \phi_1(t) / [1 + \\ &\phi_1^T(t) P_1(t-1) \phi_1(t)], \end{aligned} \quad (59)$$

$$\begin{aligned} P_1(t) &= [I - L_1(t) \phi_1^T(t)] P_1(t-1), \\ P_1(0) &= p_0 I_n, \end{aligned} \quad (60)$$

$$\begin{aligned} \hat{\theta}_i(t) &= \hat{\theta}_{i-1}(t) + \\ &L_i(t) [y_i(t) - \phi_i^T(t) \hat{\theta}_{i-1}(t)], \end{aligned} \quad (61)$$

$$\begin{aligned} L_i(t) &= P_i(t-1) \phi_i(t) / [1 + \\ &\phi_i^T(t) P_i(t-1) \phi_i(t)], \end{aligned} \quad (62)$$

$$\begin{aligned} P_i(t) &= [I - L_i(t) \phi_i^T(t)] P_i(t-1), \\ P_i(0) &= p_0 I_n, \quad i=2, 3, \dots, m. \end{aligned} \quad (63)$$

参数估计与协方差阵耦合最小二乘辨识算法(52) — (57) 只需设置两个初值 $\hat{\theta}(0) = \mathbf{1}_n / p_0$ 和 $P_m(0) = p_0 I_n$, 而参数估计耦合最小二乘辨识算法(58) — (63) 需设置 $(m+1)$ 个初值 $\hat{\theta}_m(0) = \mathbf{1}_n / p_0$ 和 $P_i(0) = p_0 I_n, i=1, 2, \dots, m$.

参数估计(全)耦合最小二乘辨识算法(58) — (63) (图3) 输出的参数估计为 $\hat{\theta}(t) = \hat{\theta}_m(t)$. 这个耦合最小二乘辨识算法的收敛性有待进一步研究.

下面通过某种“推理”(类比或假设) 给出耦合随机梯度辨识方法.

2.3 耦合随机梯度估计算法(C-SG)

梯度算法中没有协方差矩阵, 比较随机梯度辨识算法与递推最小二乘辨识算法的结构形式, 从耦合子系统最小二乘(C-SLS) 算法(39) — (41) 可以得到简单的耦合子系统随机梯度(C-SSG) 算法:

$$\begin{aligned} \hat{\theta}_i(t) &= \hat{\theta}(t-1) + \frac{\phi_i(t)}{r_i(t)} [y_i(t) - \phi_i^T(t) \hat{\theta}(t-1)], \\ \hat{\theta}(0) &= \mathbf{1}_n / p_0, \end{aligned} \quad (64)$$

$$\begin{aligned} r_i(t) &= r_i(t-1) + \|\phi_i(t)\|^2, \quad r_i(0) = 1, \\ i &= 1, 2, \dots, m, \end{aligned} \quad (65)$$

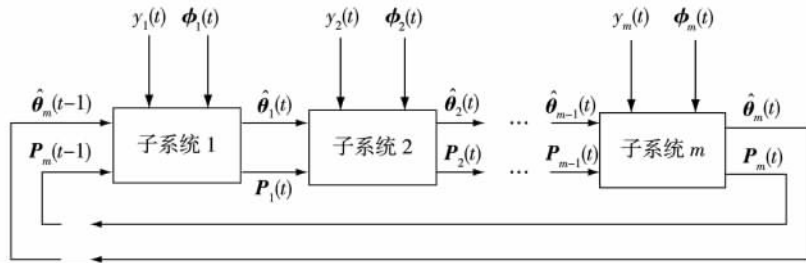


图2 耦合最小二乘算法示意

Fig. 2 The schematic diagram of the coupled least squares algorithm

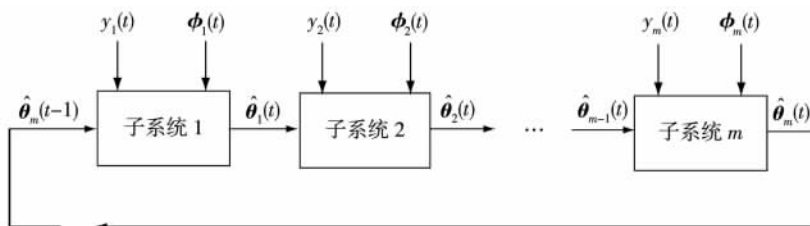


图3 参数估计耦合最小二乘辨识算法示意

Fig. 3 The schematic diagram of the parameter estimation coupled least squares algorithm

$$\hat{\theta}(t) = \frac{\hat{\theta}_1(t) + \hat{\theta}_2(t) + \dots + \hat{\theta}_m(t)}{m} \quad (66)$$

引入遗忘因子 $0 \leq \lambda \leq 1$, 可以得到遗忘因子 C-SSG 算法^[15, 27, 34]:

$$\begin{aligned} \hat{\theta}_i(t) &= \hat{\theta}_i(t-1) + \frac{\phi_i(t)}{r_i(t)} [y_i(t) - \phi_i^T(t) \hat{\theta}_i(t-1)], \\ \hat{\theta}(0) &= \mathbf{1}_n / p_0, \end{aligned} \quad (67)$$

$$\begin{aligned} r_i(t) &= \lambda r_i(t-1) + \|\phi_i(t)\|^2, \quad r_i(0) = 1, \\ 0 &\leq \lambda \leq 1, \quad i = 1, 2, \dots, m, \end{aligned} \quad (68)$$

$$\hat{\theta}(t) = \frac{\hat{\theta}_1(t) + \hat{\theta}_2(t) + \dots + \hat{\theta}_m(t)}{m} \quad (69)$$

引入收敛指数 (convergence index) ϵ , 可以得到修正 C-SSG 算法^[35-36]:

$$\begin{aligned} \hat{\theta}_i(t) &= \hat{\theta}_i(t-1) + \frac{\phi_i(t)}{r_i^\epsilon(t)} [y_i(t) - \phi_i^T(t) \hat{\theta}_i(t-1)], \\ \hat{\theta}(0) &= \mathbf{1}_n / p_0, \quad \frac{1}{2} < \epsilon \leq 1, \end{aligned} \quad (70)$$

$$\begin{aligned} r_i(t) &= r_i(t-1) + \|\phi_i(t)\|^2, \quad r_i(0) = 1, \\ i &= 1, 2, \dots, m, \end{aligned} \quad (71)$$

$$\hat{\theta}(t) = \frac{\hat{\theta}_1(t) + \hat{\theta}_2(t) + \dots + \hat{\theta}_m(t)}{m} \quad (72)$$

值得注意的是: 在递阶随机梯度辨识算法^[10, 15]中, 也可引入遗忘因子或收敛指数, 或同时引入遗忘因子和收敛指数.

定义 $r_i(t) := \text{tr}[P_i^{-1}(t)]$. 从耦合最小二乘算法 (46) — (51) 可以得到耦合随机梯度 (C-SG) 算法:

$$\begin{aligned} \hat{\theta}_i(t) &= \hat{\theta}_{i-1}(t) + \frac{\phi_i(t)}{r_i(t)} [y_i(t) - \phi_i^T(t) \hat{\theta}_{i-1}(t)], \\ \hat{\theta}_m(0) &= \mathbf{1}_n / p_0, \end{aligned} \quad (73)$$

$$r_i(t) = r_{i-1}(t) + \|\phi_i(t)\|^2, \quad i = 2, 3, \dots, m, \quad (74)$$

$$\hat{\theta}(t) = \hat{\theta}_m(t-1) + \frac{\phi_1(t)}{r_1(t)} [y_1(t) - \phi_1^T(t) \hat{\theta}_m(t-1)], \quad (75)$$

$$r_1(t) = r_m(t-1) + \|\phi_1(t)\|^2, \quad r_m(0) = 1. \quad (76)$$

也可在上述耦合随机梯度算法中引入遗忘因子或收敛指数, 得到相应的遗忘因子 C-SG 算法或修正 C-SG 算法.

耦合随机梯度算法 (73) — (76) 的子系统辨识算法间除参数估计 $\hat{\theta}_i(t)$ 耦合外, $r_i(t)$ 也是耦合的. 如果去掉 $r_i(t)$ 的耦合, 再引入遗忘因子 λ , 就得到简单的遗忘因子耦合随机梯度 (C-SG) 算法:

$$\begin{aligned} \hat{\theta}_i(t) &= \hat{\theta}_i(t-1) + \frac{\phi_i(t)}{r_i(t)} [y_i(t) - \phi_i^T(t) \hat{\theta}_i(t-1)], \\ \hat{\theta}_m(0) &= \mathbf{1}_n / p_0, \end{aligned} \quad (77)$$

$$r_1(t) = \lambda r_1(t-1) + \|\phi_1(t)\|^2, \quad 0 \leq \lambda \leq 1,$$

$$r_1(0) = 1, \quad (78)$$

$$\hat{\theta}_i(t) = \hat{\theta}_{i-1}(t) + \frac{\phi_i(t)}{r_i(t)} [y_i(t) - \phi_i^T(t) \hat{\theta}_{i-1}(t)], \quad (79)$$

$$\begin{aligned} r_i(t) &= \lambda r_i(t-1) + \|\phi_i(t)\|^2, \quad r_i(0) = 1, \\ i &= 2, 3, \dots, m. \end{aligned} \quad (80)$$

当然, 也可以引入多新息辨识理论, 进一步推导出多变量系统的耦合多新息随机梯度辨识算法, 耦合多新息最小二乘辨识方法. 这些辨识算法的收敛性有待进一步研究.

3 部分耦合随机梯度辨识方法 (PC-SG)

考虑状态空间模型描述的线性多变量系统:

$$\begin{cases} \mathbf{x}(t+1) = \mathbf{A}\mathbf{x}(t) + \mathbf{B}\mathbf{u}(t), \\ \mathbf{y}(t) = \mathbf{C}\mathbf{x}(t). \end{cases} \quad (81)$$

其中 $\mathbf{x}(t) \in \mathbf{R}^n$ 为状态向量, $\mathbf{u}(t) = [u_1(t), u_2(t), \dots, u_r(t)]^T \in \mathbf{R}^r$ 为系统输入向量, $\mathbf{y}(t) = [y_1(t), y_2(t), \dots, y_m(t)]^T \in \mathbf{R}^m$ 为系统输出向量, $\mathbf{A}, \mathbf{B}, \mathbf{C}$ 为适当维数的系统矩阵.

状态空间模型 (81) 的输入输出与其传递函数矩阵 $\mathbf{G}(z) \in \mathbf{R}^{m \times r}$ 有下列关系:

$$\mathbf{y}(t) = \mathbf{G}(z) \mathbf{u}(t), \quad (82)$$

$$\mathbf{G}(z) = \mathbf{C}(z\mathbf{I} - \mathbf{A})^{-1}\mathbf{B} = \frac{\mathbf{C} \text{adj}[z\mathbf{I} - \mathbf{A}]\mathbf{B}}{\det[z\mathbf{I} - \mathbf{A}]}$$

$$= \frac{z^{-n} \mathbf{C} \text{adj}[z\mathbf{I} - \mathbf{A}]\mathbf{B}}{z^{-n} \det[z\mathbf{I} - \mathbf{A}]} =: \frac{\mathbf{Q}(z)}{\alpha(z)}.$$

其中 $\alpha(z)$ 是系统的 n 次特征多项式, 定义为传递矩阵的各元最小公分母首一多项式, 即 $\mathbf{G}(z)$ 的最小公分母 $\mathbf{Q}(z)$ 是 z^{-1} 的矩阵多项式, 它们可以表达为

$$\alpha(z) := z^{-n} \det[z\mathbf{I} - \mathbf{A}] =$$

$$1 + \alpha_1 z^{-1} + \alpha_2 z^{-2} + \dots + \alpha_n z^{-n}, \quad \alpha_i \in \mathbf{R},$$

$$\mathbf{Q}(z) := z^{-n} \mathbf{C} \text{adj}[z\mathbf{I} - \mathbf{A}]\mathbf{B} =$$

$$\mathbf{Q}_1 z^{-1} + \mathbf{Q}_2 z^{-2} + \dots + \mathbf{Q}_n z^{-n}, \quad \mathbf{Q}_i \in \mathbf{R}^{m \times r}.$$

特征多项式 $\alpha(z)$ 的辨识, 对于控制系统极点配置、观测器设计、稳定性分析十分重要.

式 (82) 可以写作 $\mathbf{y}(t) = \frac{\mathbf{Q}(z)}{\alpha(z)} \mathbf{u}(t)$, 将之称为多

变量系统传递函数阵主模型 (main model), 其特征是传递函数阵 $\mathbf{G}(z) = \frac{\mathbf{Q}(z)}{\alpha(z)}$ 有一个系统特征多项式

作为公分母, 分子为一个矩阵多项式. 主模型是相对于分子系统的子模型 (submodel) 而言的. 上式又可写为 $\alpha(z) \mathbf{y}(t) = \mathbf{Q}(z) \mathbf{u}(t)$. 基于这个模型, 引入一个随机干扰项 $\mathbf{v}(t) \in \mathbf{R}^m$, 可得类多变量 CAR 系统 (类多变量 ARX 系统), 简称多变量 CAR-like 系统

(多变量 ARX-like 系统)^[15-16, 37]:

$$\alpha(z) y(t) = Q(z) u(t) + v(t).$$

其中 $v(t) = [v_1(t) \ v_2(t) \ \dots \ v_m(t)]^T \in \mathbf{R}^m$ 是一个零均值随机噪声向量. 将移位算子 z^{-1} 的多项式 $\alpha(z)$ 和多项式矩阵 $Q(z)$ 代入上式得到

$$y(t) + \alpha_1 y(t-1) + \alpha_2 y(t-2) + \dots + \alpha_n y(t-n) = Q_1 u(t-1) + Q_2 u(t-2) + \dots + Q_n u(t-n) + v(t).$$

进一步可以写为

$$y(t) + [y(t-1) \ y(t-2) \ \dots \ y(t-n)] \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_n \end{bmatrix} = [Q_1 \ Q_2 \ \dots \ Q_n] \begin{bmatrix} u(t-1) \\ u(t-2) \\ \vdots \\ u(t-n) \end{bmatrix} + v(t). \quad (83)$$

定义参数矩阵 θ 、参数向量 α 、输入信息向量 $\varphi(t)$ 和输出信息矩阵 $\psi(t)$ 如下:

$$\theta^T := [Q_1 \ Q_2 \ \dots \ Q_n] \in \mathbf{R}^{m \times (nr)},$$

$$\alpha := \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_n \end{bmatrix} \in \mathbf{R}^n, \quad \varphi(t) := \begin{bmatrix} u(t-1) \\ u(t-2) \\ \vdots \\ u(t-n) \end{bmatrix} \in \mathbf{R}^{nr},$$

$$\psi(t) := [y(t-1) \ y(t-2) \ \dots \ y(t-n)] \in \mathbf{R}^{m \times n}.$$

于是, 从式(83)可以得到多变量 ARX-like 系统的递阶辨识模型^[15-16]:

$$y(t) + \psi(t) \alpha = \theta^T \varphi(t) + v(t). \quad (84)$$

3.1 随机梯度辨识算法(SG)

辨识模型(84)包含一个参数向量 $\alpha \in \mathbf{R}^n$ 和一个参数矩阵 $\theta \in \mathbf{R}^{(nr) \times m}$. 为了辨识 α 和 θ , 一种方法是将模型(84)转换为一种新的形式. 令 $\theta_i \in \mathbf{R}^{nr}$ 为 θ 的第 i 列, 置

$$\vartheta := \begin{bmatrix} \alpha \\ \text{col}[\theta] \end{bmatrix} = \begin{bmatrix} \alpha \\ \theta_1 \\ \theta_2 \\ \vdots \\ \theta_m \end{bmatrix} \in \mathbf{R}^{n+mnr},$$

$$\Phi(t) := [-\psi(t) \ I_m \otimes \varphi^T(t)] \in \mathbf{R}^{m \times (n+mnr)}. \quad (85)$$

式(84)可以写为

$$y(t) = \Phi(t) \vartheta + v(t). \quad (86)$$

这里参数向量 ϑ 包含了系统的所有参数.

令 $\hat{\alpha}(t)$ 、 $\hat{\theta}(t)$ 和 $\hat{\vartheta}(t)$ 分别代表 α 、 θ 和 ϑ 在时

刻 t 的估计. 使用负梯度搜索, 极小化准则函数

$$J_3(\vartheta) := \|y(t) - \Phi(t) \vartheta\|^2,$$

可以得到估计参数向量 ϑ 的随机梯度算法:

$$\hat{\vartheta}(t) = \hat{\vartheta}(t-1) + \frac{\Phi^T(t)}{R(t)} [y(t) - \Phi(t) \hat{\vartheta}(t-1)], \quad (87)$$

$$R(t) = R(t-1) + \|\Phi(t)\|^2, \quad R(0) = 1. \quad (88)$$

尽管这个随机梯度算法(87) — (88) 能够估计参数向量 ϑ , 但计算量大, 原因在于辨识模型(86)中信息矩阵 $\Phi(t)$ 包含了太多的零元, 导致参数向量维数太大. 令 $\psi_i(t) \in \mathbf{R}^n$ 为 $\psi(t)$ 的第 i 行, 即

$$\psi(t) := \begin{bmatrix} \psi_1(t) \\ \psi_2(t) \\ \vdots \\ \psi_m(t) \end{bmatrix} \in \mathbf{R}^{m \times n},$$

那么矩阵 $\Phi(t)$ 可以展开为

$$\Phi(t) = [-\psi(t) \ I_m \otimes \varphi^T(t)] = \begin{bmatrix} -\psi_1(t) & \varphi^T(t) & \mathbf{0} & \dots & \mathbf{0} \\ -\psi_2(t) & \mathbf{0} & \varphi^T(t) & \dots & \mathbf{0} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ -\psi_m(t) & \mathbf{0} & \mathbf{0} & \dots & \varphi^T(t) \end{bmatrix} \in \mathbf{R}^{m \times (n+mnr)}. \quad (89)$$

从这里可以看出: 矩阵 $\Phi(t)$ 包含了 $mn(mr+1)$ 元, 其中有 $(m-1)mnr$ 个零元. 同样, 基于辨识模型(86)的最小二乘辨识方法计算量也很大.

下面推导计算量小的子系统随机梯度算法和部分耦合随机梯度算法.

3.2 子系统随机梯度辨识算法(SSG)

定义子系统信息向量

$$\phi_i(t) := \begin{bmatrix} -\psi_i^T(t) \\ \varphi(t) \end{bmatrix} \in \mathbf{R}^{n+nr}. \quad (90)$$

根据式(89)中 $\Phi(t)$ 的结构, 式(86)可以写为

$$\begin{bmatrix} y_1(t) \\ y_2(t) \\ \vdots \\ y_m(t) \end{bmatrix} = \begin{bmatrix} -\psi_1(t) & \varphi^T(t) & \mathbf{0} & \dots & \mathbf{0} \\ -\psi_2(t) & \mathbf{0} & \varphi^T(t) & \dots & \mathbf{0} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ -\psi_m(t) & \mathbf{0} & \mathbf{0} & \dots & \varphi^T(t) \end{bmatrix} \begin{bmatrix} \alpha \\ \theta_1 \\ \theta_2 \\ \vdots \\ \theta_m \end{bmatrix} + \begin{bmatrix} v_1(t) \\ v_2(t) \\ \vdots \\ v_m(t) \end{bmatrix}.$$

其第 i 个子系统为(可以分解为下列 m 个子系统)

$$y_i(t) = [-\psi_i(t) \ \varphi^T(t)] \begin{bmatrix} \alpha \\ \theta_i \end{bmatrix} + v_i(t),$$

$$= \phi_i^T(t) \begin{bmatrix} \alpha \\ \theta_i \end{bmatrix} + v_i(t), \quad i=1, 2, \dots, m. \quad (91)$$

使用负梯度搜索 极小化准则函数

$$J_A(\alpha, \theta_i) = \left\{ y_i(t) - \phi_i^T(t) \begin{bmatrix} \alpha \\ \theta_i \end{bmatrix} \right\}^2, \quad i=1, 2, \dots, m.$$

可以得到估计参数向量 α 和 θ_i 的子系统随机梯度 (SSG) 算法:

$$\begin{bmatrix} \hat{\alpha}(t) \\ \hat{\theta}_i(t) \end{bmatrix} = \begin{bmatrix} \hat{\alpha}(t-1) \\ \hat{\theta}_i(t-1) \end{bmatrix} + \frac{\phi_i(t)}{R_i(t)} \left\{ y_i(t-1) - \phi_i^T(t-1) \begin{bmatrix} \hat{\alpha}(t-1) \\ \hat{\theta}_i(t-1) \end{bmatrix} \right\}, \quad (92)$$

$$R_i(t) = R_i(t-1) + \|\phi_i(t)\|^2, \quad R_i(0) = 1, \quad i=1, 2, \dots, m. \quad (93)$$

SSG 算法 (92) — (93) 可以给出参数向量 α 和 θ_i 的估计 $i=1, 2, \dots, m$, 但是对于每个 i α 都估计了一次, 为区分每个子系统中 α 的估计, 加下标 i , 记作 $\hat{\alpha}_i(t)$, 便得到下列 SSG 辨识算法:

$$\begin{bmatrix} \hat{\alpha}(t) \\ \hat{\theta}_i(t) \end{bmatrix} = \begin{bmatrix} \hat{\alpha}(t-1) \\ \hat{\theta}_i(t-1) \end{bmatrix} + \frac{\phi_i(t)}{R_i(t)} \left\{ y_i(t-1) - \phi_i^T(t-1) \begin{bmatrix} \hat{\alpha}(t-1) \\ \hat{\theta}_i(t-1) \end{bmatrix} \right\}, \quad \hat{\alpha}_i(0) = \mathbf{1}_n/p_0, \quad \hat{\theta}_i(0) = \mathbf{1}_m/p_0, \quad (94)$$

$$R_i(t) = R_i(t-1) + \|\phi_i(t)\|^2, \quad R_i(0) = 1, \quad i=1, 2, \dots, m. \quad (95)$$

子系统随机梯度辨识算法 (94) — (95) 方框如图 4 所示.

注 3 式 (94) 中 $\hat{\theta}_i(t)$ 是参数向量 θ_i 的估计, $\hat{\alpha}_i(t)$ 是子系统间共同参数向量 α 的估计. 同理, 将每个子系统的估计 $\hat{\alpha}_i(t)$ 的平均值作为参数向量 α 的估计, 即

$$\hat{\alpha}(t) = \frac{\hat{\alpha}_1(t) + \hat{\alpha}_2(t) + \dots + \hat{\alpha}_m(t)}{m} \in \mathbf{R}^n. \quad (96)$$

用平均值 $\hat{\alpha}(t-1)$ 代替 SSG 算法 (94) — (95) 中的 $\hat{\alpha}_i(t-1)$, 就得到一个简单的部分耦合子系统随机梯度 (PC-SSG) 算法:

$$\begin{bmatrix} \hat{\alpha}(t) \\ \hat{\theta}_i(t) \end{bmatrix} = \begin{bmatrix} \hat{\alpha}(t-1) \\ \hat{\theta}_i(t-1) \end{bmatrix} + \frac{\phi_i(t)}{R_i(t)} \left\{ y_i(t-1) - \phi_i^T(t-1) \begin{bmatrix} \hat{\alpha}(t-1) \\ \hat{\theta}_i(t-1) \end{bmatrix} \right\}, \quad \hat{\alpha}(0) = \mathbf{1}_n/p_0, \quad \hat{\theta}_i(0) = \mathbf{1}_m/p_0, \quad (97)$$

$$R_i(t) = R_i(t-1) + \|\phi_i(t)\|^2, \quad R_i(0) = 1,$$

$$i=1, 2, \dots, m, \quad (98)$$

$$\hat{\alpha}(t) = \frac{\hat{\alpha}_1(t) + \hat{\alpha}_2(t) + \dots + \hat{\alpha}_m(t)}{m}. \quad (99)$$

下面讨论部分耦合随机梯度辨识算法, 以避免参数向量 α 的冗余估计.

3.3 部分耦合随机梯度辨识算法 (PC-SG)

在子系统辨识模型 (91) 中, 只有参数向量 $\alpha \in \mathbf{R}^n$ 是耦合的, 即各子系统有一个共同的参数向量 α , 借助于前面讨论的全耦合最小二乘辨识方法 (即耦合最小二乘算法) 的推导思路, 不难得到部分耦合随机梯度辨识算法 (Partially Coupled Stochastic Gradient PC-SG) [13]:

$$\begin{bmatrix} \hat{\alpha}(t) \\ \hat{\theta}_i(t) \end{bmatrix} = \begin{bmatrix} \hat{\alpha}(t-1) \\ \hat{\theta}_i(t-1) \end{bmatrix} + \frac{\phi_i(t)}{R_i(t)} \left\{ y_i(t-1) - \phi_i^T(t-1) \begin{bmatrix} \hat{\alpha}(t-1) \\ \hat{\theta}_i(t-1) \end{bmatrix} \right\}, \quad \hat{\theta}_i(0) = \mathbf{1}_m/p_0, \quad (100)$$

$$R_i(t) = R_i(t-1) + \|\phi_i(t)\|^2, \quad R_i(0) = 1, \quad i=2, 3, \dots, m, \quad (101)$$

$$\begin{bmatrix} \hat{\alpha}(t) \\ \hat{\theta}_i(t) \end{bmatrix} = \begin{bmatrix} \hat{\alpha}_m(t-1) \\ \hat{\theta}_i(t-1) \end{bmatrix} + \frac{\phi_i(t)}{R_i(t)} \left\{ y_i(t-1) - \phi_i^T(t-1) \begin{bmatrix} \hat{\alpha}_m(t-1) \\ \hat{\theta}_i(t-1) \end{bmatrix} \right\}, \quad \hat{\alpha}_m(0) = \mathbf{1}_n/p_0, \quad \hat{\theta}_i(0) = \mathbf{1}_m/p_0, \quad (102)$$

$$R_1(t) = R_1(t-1) + \|\phi_1(t)\|^2, \quad R_1(0) = 1, \quad (103)$$

$$\varphi(t) = [\mathbf{u}^T(t-1) \quad \mathbf{u}^T(t-2) \quad \dots \quad \mathbf{u}^T(t-n)]^T, \quad (104)$$

$$\psi(t) = [y(t-1) \quad y(t-2) \quad \dots \quad y(t-n)], \quad (105)$$

$$\psi(t) = \begin{bmatrix} \psi_1(t) \\ \psi_2(t) \\ \vdots \\ \psi_m(t) \end{bmatrix}, \quad \phi_i(t) = \begin{bmatrix} -\psi_i^T(t) \\ \varphi(t) \end{bmatrix}. \quad (106)$$

部分耦合随机梯度辨识方法 (100) — (106) 的方框如图 5 所示, 其中只有参数估计向量 $\hat{\alpha}_i(t)$ 是耦合在各子系统辨识算法之间, 参数向量 $\hat{\theta}_i(t)$ 是没有耦合的, 它是各子系统的参数估计. 这就是部分耦合名称的来历.

参数向量 $\alpha, \theta_1, \theta_2, \dots, \theta_m$ 的估计分别是

$$\hat{\alpha}(t) : = \hat{\alpha}_m(t), \quad \hat{\theta}_1(t), \quad \hat{\theta}_2(t), \quad \dots, \quad \hat{\theta}_m(t).$$

注 4 与 SG 算法 (87) — (88) 和 PC-SSG 算法 (97) — (99) 比, PC-SG 算法 (100) — (106) 由于避免了参数向量 α 的冗余估计, 具有高的计算效率. 它们每一步的计算量比较如表 2 所示. 当 $n=5$ 时, 一个 5

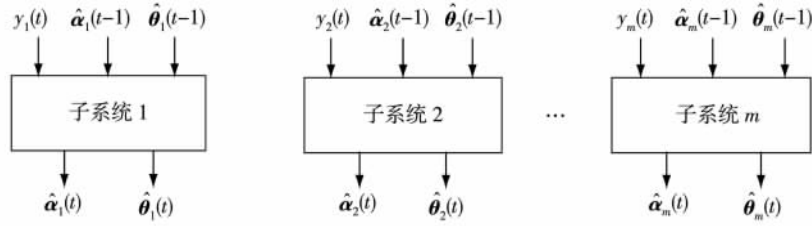


图 4 子系统随机梯度算法示意

Fig. 4 The schematic diagram of the subsystem stochastic gradient algorithm

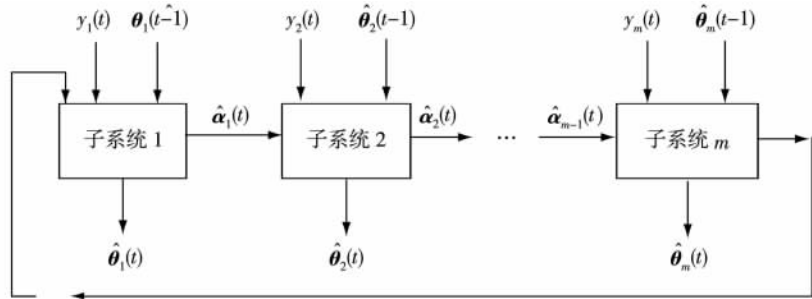


图 5 分耦合随机梯度算法示意

Fig. 5 The schematic diagram of the partially coupled stochastic gradient algorithm

个输入 5 个输出的 5 阶系统,其每一步的乘法次数和加法次数参见表 2 中数据.可见,PC-SG 算法计算量最小.

表 2 辨识算法计算量比较

Table 2 The comparison of computational efficiency

算法	乘法次数	加法次数
SG	$3m(n + mnr) + m = 1\ 955$	$3m(n + mnr) = 1\ 950$
PC-SSG	$3m(n + nr) + m + n = 460$	$m(n + nr) + (m - 1)n = 470$
PC-SG	$3m(n + nr) + m = 455$	$3m(n + nr) = 450$

一般来说,一个递推算法的计算量可以用其每一步的乘法次数和加法次数来衡量.有的文献上把一次乘法次数或一次加法次数称为一个 flop^[32].因此,每递推计算一步,SG 算法需要 $1\ 955 + 1\ 950 = 3\ 905$ flop; PC-SSG 算法需要 $460 + 470 = 930$ flop; PC-SG 算法需要 $455 + 450 = 905$ flop.

利用 PC-SG 算法(100) — (106) 计算参数估计 $\hat{\alpha}_m(t)$ 和 $\hat{\theta}_i(t)$ ($i = 1, 2, \dots, m$) 的步骤如下.

1) 设置初值. $t = 1$, $\hat{\alpha}_m(0) = \mathbf{1}_n/p_0$, $\hat{\theta}_i(0) = \mathbf{1}_{n_0}/p_0$, $R_i(0) = 1$, $i = 1, 2, \dots, m$, $p_0 = 10^6$.

2) 收集输入输出数据 $u(t)$ 和 $y(t)$,由式(104) 构成输入信息向量 $\varphi(t)$,由式(105) 构成输出信息矩阵 $\Psi(t)$,由式(106) 抽出 $\psi_i(t)$ ($i = 1, 2, \dots, m$),

并构成 $\phi_i(t)$.

3) 由式(103) 计算 $R_1(t)$,由式(102) 刷新参数估计 $\hat{\alpha}_1(t)$ 和 $\hat{\theta}_1(t)$.

4) 对于每一个 i ($i = 2, 3, \dots, m$),由式(101) 计算 $R_i(t)$,由式(100) 刷新参数估计 $\hat{\alpha}_i(t)$ 和 $\hat{\theta}_i(t)$.

5) t 增 1 转到第 2 步.

利用 PC-SG 算法(100) — (106) 计算参数估计 $\hat{\alpha}_m(t)$ 和 $\hat{\theta}_i(t)$ ($i = 1, 2, \dots, m$) 的流程如图 6 所示.

为了提高参数估计收敛速度,可以在随机梯度算法中引入遗忘因子 λ ,即式(88), (101) 和(103) 中引入 λ ,得到

$$R(t) = \lambda R(t-1) + \|\Phi(t)\|^2, \quad R(0) = 1, \quad 0 \leq \lambda \leq 1, \quad (107)$$

$$R_i(t) = \lambda R_i(t-1) + \|\varphi_i(t)\|^2, \quad R_i(0) = 1, \quad i = 1, 2, \dots, m, \quad 0 \leq \lambda \leq 1, \quad (108)$$

就获得带遗忘因子 λ 的 SG 算法(87) 和(107),以及带遗忘因子 λ 的 PC-SG 算法(100), (102), (108), (104) — (106).当然,也可以引入收敛指数来提高随机梯度算法的收敛速度.

4 部分耦合最小二乘辨识方法(PC-LS)

本节介绍几个部分耦合最小二乘辨识方法,部分内容引自笔者指导的博士生刘艳君发表在 2012

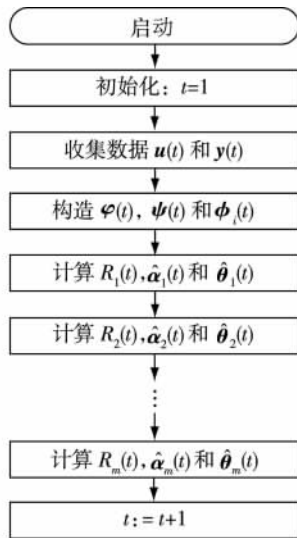


图 6 计算参数估计参数估计 $\hat{\alpha}_m(t)$ 和 $\hat{\theta}_i(t)$ 流程 ($i=1, 2, \dots, m$)

Fig. 6 The flowchart of computing the parameter estimates $\hat{\alpha}_m(t)$ and $\hat{\theta}_i(t)$ ($i=1, 2, \dots, m$)

年中国控制与决策会议的论文^[38].

4.1 最小二乘辨识算法 (LS)

对于辨识模型 (86) 极小化准则函数

$$J(\boldsymbol{\vartheta}) = \sum_{j=1}^l \|y(j) - \boldsymbol{\Phi}(j)\boldsymbol{\vartheta}\|^2$$

得到下列多变量递推最小二乘辨识算法:

$$\hat{\boldsymbol{\vartheta}}(t) = \hat{\boldsymbol{\vartheta}}(t-1) + L(t) [y(t) - \boldsymbol{\Phi}(t)\hat{\boldsymbol{\vartheta}}(t-1)], \quad (109)$$

$$L(t) = P(t-1)\boldsymbol{\Phi}^T(t) [\mathbf{I}_m + \boldsymbol{\Phi}(t)P(t-1)\boldsymbol{\Phi}^T(t)]^{-1}, \quad (110)$$

$$P(t) = [\mathbf{I}_{n+mnr} - L(t)\boldsymbol{\Phi}(t)]P(t-1),$$

$$P(0) = p_0\mathbf{I}_{n+mnr}. \quad (111)$$

这里 $\hat{\boldsymbol{\vartheta}}(t)$ 是时刻 t 参数向量 $\boldsymbol{\vartheta}$ 的估计 $P(t)$ 是 $(n + mnr) \times (n + mnr)$ 维协方差阵.

注 5 多变量 RLS 算法 (109) — (111) 在每一步递推计算中都要计算大维数协方差阵 $P(t)$, 导致算法的计算量很大.

4.2 子系统最小二乘辨识算法 (SLS)

对于辨识模型 (86) 极小化准则函数

$$J(\boldsymbol{\alpha}, \boldsymbol{\theta}_i) = \sum_{j=1}^l \left\{ y_i(j) - \boldsymbol{\phi}_i^T(j) \begin{bmatrix} \boldsymbol{\alpha} \\ \boldsymbol{\theta}_i \end{bmatrix} \right\}^2, \quad i=1, 2, \dots, m.$$

可以得到估计参数向量 $\boldsymbol{\alpha}$ 和 $\boldsymbol{\theta}_i$ 的子系统最小二乘 (SLS) 算法:

$$\begin{bmatrix} \hat{\boldsymbol{\alpha}}(t) \\ \hat{\boldsymbol{\theta}}_i(t) \end{bmatrix} = \begin{bmatrix} \hat{\boldsymbol{\alpha}}(t-1) \\ \hat{\boldsymbol{\theta}}_i(t-1) \end{bmatrix} +$$

$$P_i(t)\boldsymbol{\varphi}_i(t) \left\{ y_i(t) - \boldsymbol{\phi}_i^T(t) \begin{bmatrix} \hat{\boldsymbol{\alpha}}(t-1) \\ \hat{\boldsymbol{\theta}}_i(t-1) \end{bmatrix} \right\}, \quad (112)$$

$$P_i^{-1}(t) = P_i^{-1}(t-1) + \boldsymbol{\phi}_i(t)\boldsymbol{\phi}_i^T(t), \quad i=1, 2, \dots, m. \quad (113)$$

出于同样的原因, 为了区别每个子系统对共同参数向量 $\boldsymbol{\alpha}$ 的估计, 在子系统的参数估计 $\hat{\boldsymbol{\alpha}}(t)$ 上加上下标 i , 得到

$$\begin{bmatrix} \hat{\boldsymbol{\alpha}}_i(t) \\ \hat{\boldsymbol{\theta}}_i(t) \end{bmatrix} = \begin{bmatrix} \hat{\boldsymbol{\alpha}}_i(t-1) \\ \hat{\boldsymbol{\theta}}_i(t-1) \end{bmatrix} +$$

$$P_i(t)\boldsymbol{\phi}_i(t) \left\{ y_i(t) - \boldsymbol{\phi}_i^T(t) \begin{bmatrix} \hat{\boldsymbol{\alpha}}_i(t-1) \\ \hat{\boldsymbol{\theta}}_i(t-1) \end{bmatrix} \right\}. \quad (114)$$

对于 $\boldsymbol{\alpha}$ 的冗余估计, 用它们的平均值作为 $\boldsymbol{\alpha}$ 的估计, 即

$$\hat{\boldsymbol{\alpha}}(t) = \frac{\hat{\boldsymbol{\alpha}}_1(t) + \hat{\boldsymbol{\alpha}}_2(t) + \dots + \hat{\boldsymbol{\alpha}}_m(t)}{m}. \quad (115)$$

式 (113) — (114) 构成了子系统最小二乘 (SLS) 辨识算法, 总结如下:

$$\begin{bmatrix} \hat{\boldsymbol{\alpha}}_i(t) \\ \hat{\boldsymbol{\theta}}_i(t) \end{bmatrix} = \begin{bmatrix} \hat{\boldsymbol{\alpha}}_i(t-1) \\ \hat{\boldsymbol{\theta}}_i(t-1) \end{bmatrix} +$$

$$P_i(t)\boldsymbol{\varphi}_i(t) \left\{ y_i(t) - \boldsymbol{\phi}_i^T(t) \begin{bmatrix} \hat{\boldsymbol{\alpha}}_i(t-1) \\ \hat{\boldsymbol{\theta}}_i(t-1) \end{bmatrix} \right\}, \quad (116)$$

$$P_i^{-1}(t) = P_i^{-1}(t-1) + \boldsymbol{\phi}_i(t)\boldsymbol{\phi}_i^T(t),$$

$$P(0) = p_0\mathbf{I}_{n+nr}, \quad i=1, 2, \dots, m, \quad (117)$$

$$\hat{\boldsymbol{\alpha}}(t) = \frac{\hat{\boldsymbol{\alpha}}_1(t) + \hat{\boldsymbol{\alpha}}_2(t) + \dots + \hat{\boldsymbol{\alpha}}_m(t)}{m},$$

$$\hat{\boldsymbol{\alpha}}(0) = \mathbf{1}_n/p_0, \quad \hat{\boldsymbol{\theta}}_i(0) = \mathbf{1}_{nr}/p_0. \quad (118)$$

应用矩阵求逆引理, 可以等价写为

$$\begin{bmatrix} \hat{\boldsymbol{\alpha}}_i(t) \\ \hat{\boldsymbol{\theta}}_i(t) \end{bmatrix} = \begin{bmatrix} \hat{\boldsymbol{\alpha}}_i(t-1) \\ \hat{\boldsymbol{\theta}}_i(t-1) \end{bmatrix} +$$

$$L_i(t) \left\{ y_i(t) - \boldsymbol{\phi}_i^T(t) \begin{bmatrix} \hat{\boldsymbol{\alpha}}_i(t-1) \\ \hat{\boldsymbol{\theta}}_i(t-1) \end{bmatrix} \right\}, \quad (119)$$

$$L_i(t) = P_i(t)\boldsymbol{\phi}_i(t) =$$

$$P_i(t-1)\boldsymbol{\phi}_i(t) [1 + \boldsymbol{\phi}_i^T(t)P_i(t-1)\boldsymbol{\phi}_i(t)]^{-1}, \quad (120)$$

$$P_i(t) = [\mathbf{I}_{n+nr} - L_i(t)\boldsymbol{\phi}_i^T(t)]P_i(t-1),$$

$$P(0) = p_0\mathbf{I}_{n+nr}, \quad i=1, 2, \dots, m, \quad (121)$$

$$\hat{\boldsymbol{\alpha}}(t) = \frac{\hat{\boldsymbol{\alpha}}_1(t) + \hat{\boldsymbol{\alpha}}_2(t) + \dots + \hat{\boldsymbol{\alpha}}_m(t)}{m},$$

$$\hat{\boldsymbol{\alpha}}(0) = \mathbf{1}_n/p_0, \quad \hat{\boldsymbol{\theta}}_i(0) = \mathbf{1}_{nr}/p_0. \quad (122)$$

4.3 部分耦合子系统最小二乘辨识算法 (PC-SLS)

用平均值 $\hat{\boldsymbol{\alpha}}(t-1)$ 代替 SLS 算法 (116) — (118) 或 (119) — (122) 中的 $\hat{\boldsymbol{\alpha}}_i(t-1)$, 就得到一个简单的部分耦合子系统最小二乘 (PC-SLS) 算法:

$$\begin{bmatrix} \hat{\alpha}_i(t) \\ \hat{\theta}_i(t) \end{bmatrix} = \begin{bmatrix} \hat{\alpha}(t-1) \\ \hat{\theta}_i(t-1) \end{bmatrix} + P_i(t) \phi_i(t) \left\{ y_i(t) - \phi_i^T(t) \begin{bmatrix} \hat{\alpha}(t-1) \\ \hat{\theta}_i(t-1) \end{bmatrix} \right\}, \quad (123)$$

$$\begin{aligned} P_i^{-1}(t) &= P_i^{-1}(t-1) + \phi_i(t) \phi_i^T(t), \\ P(0) &= p_0 \mathbf{I}_{n+nr}, \quad i=1, 2, \dots, m, \end{aligned} \quad (124)$$

$$\begin{aligned} \hat{\alpha}(t) &= \frac{\hat{\alpha}_1(t) + \hat{\alpha}_2(t) + \dots + \hat{\alpha}_m(t)}{m}, \\ \hat{\alpha}(0) &= \mathbf{1}_n/p_0, \quad \hat{\theta}_i(0) = \mathbf{1}_{nr}/p_0. \end{aligned} \quad (125)$$

应用矩阵求逆引理, 可以等价写为

$$\begin{bmatrix} \hat{\alpha}_i(t) \\ \hat{\theta}_i(t) \end{bmatrix} = \begin{bmatrix} \hat{\alpha}(t-1) \\ \hat{\theta}_i(t-1) \end{bmatrix} + L_i(t) \left\{ y_i(t) - \phi_i^T(t) \begin{bmatrix} \hat{\alpha}(t-1) \\ \hat{\theta}_i(t-1) \end{bmatrix} \right\}, \quad (126)$$

$$\begin{aligned} L_i(t) &= P_i(t) \phi_i(t) = \\ &P_i(t-1) \phi_i(t) [1 + \phi_i^T(t) P_i(t-1) \phi_i(t)]^{-1}, \end{aligned} \quad (127)$$

$$\begin{aligned} P_i(t) &= [\mathbf{I}_{n+nr} - L_i(t) \phi_i^T(t)] P_i(t-1), \\ P(0) &= p_0 \mathbf{I}_{n+nr}, \quad i=1, 2, \dots, m, \end{aligned} \quad (128)$$

$$\begin{aligned} \hat{\alpha}(t) &= \frac{\hat{\alpha}_1(t) + \hat{\alpha}_2(t) + \dots + \hat{\alpha}_m(t)}{m}, \\ \hat{\alpha}(0) &= \mathbf{1}_n/p_0, \quad \hat{\theta}_i(0) = \mathbf{1}_{nr}/p_0. \end{aligned} \quad (129)$$

这个 PC-SLS 算法的特点是: 各子系统协方差阵 $P_i(t) \in \mathbf{R}^{(nr) \times (nr)}$ 是独立的.

4.4 部分耦合最小二乘辨识算法(PC-LS)

利用耦合辨识概念, 仿照 C-LS 算法的推导, 不难得到下列部分耦合最小二乘辨识算法 (Partially Coupled Least Squares identification algorithm, PC-LS)^[38]:

$$\begin{bmatrix} \hat{\alpha}_1(t) \\ \hat{\theta}_1(t) \end{bmatrix} = \begin{bmatrix} \hat{\alpha}_m(t-1) \\ \hat{\theta}_1(t-1) \end{bmatrix} + P_1(t) \phi_1(t) \left\{ y_1(t) - \phi_1^T(t) \begin{bmatrix} \hat{\alpha}_m(t-1) \\ \hat{\theta}_1(t-1) \end{bmatrix} \right\}, \quad (130)$$

$$\begin{aligned} P_1^{-1}(t) &= P_m^{-1}(t-1) + \phi_1(t) \phi_1^T(t), \\ P_m(0) &= p_0 \mathbf{I}_{n+nr}, \quad \hat{\alpha}_m(0) = \mathbf{1}_n/p_0, \end{aligned} \quad (131)$$

$$\begin{bmatrix} \hat{\alpha}_i(t) \\ \hat{\theta}_i(t) \end{bmatrix} = \begin{bmatrix} \hat{\alpha}_{i-1}(t) \\ \hat{\theta}_i(t-1) \end{bmatrix} + P_i(t) \phi_i(t) \left\{ y_i(t) - \phi_i^T(t) \begin{bmatrix} \hat{\alpha}_{i-1}(t) \\ \hat{\theta}_i(t-1) \end{bmatrix} \right\}, \quad (132)$$

$$\begin{aligned} P_i^{-1}(t) &= P_{i-1}^{-1}(t) + \phi_i(t) \phi_i^T(t), \\ \hat{\theta}_i(0) &= \mathbf{1}_{nr}/p_0, \quad i=2, 3, \dots, m. \end{aligned} \quad (133)$$

应用矩阵求逆引理, PC-LS 算法总结如下:

$$\begin{bmatrix} \hat{\alpha}_1(t) \\ \hat{\theta}_1(t) \end{bmatrix} = \begin{bmatrix} \hat{\alpha}_m(t-1) \\ \hat{\theta}_1(t-1) \end{bmatrix} + L_1(t) \left\{ y_1(t) - \phi_1^T(t) \begin{bmatrix} \hat{\alpha}_m(t-1) \\ \hat{\theta}_1(t-1) \end{bmatrix} \right\}, \quad (134)$$

$$L_1(t) = \frac{P_m(t-1) \phi_1(t)}{1 + \phi_1^T(t) P_m(t-1) \phi_1(t)}, \quad (135)$$

$$\begin{aligned} P_1(t) &= [\mathbf{I} - L_1(t) \phi_1^T(t)] P_m(t-1), \\ P_m(0) &= p_0 \mathbf{I}_{n+nr}, \quad \hat{\alpha}_m(0) = \mathbf{1}_n/p_0, \end{aligned} \quad (136)$$

$$\begin{bmatrix} \hat{\alpha}_i(t) \\ \hat{\theta}_i(t) \end{bmatrix} = \begin{bmatrix} \hat{\alpha}_{i-1}(t) \\ \hat{\theta}_i(t-1) \end{bmatrix} + L_i(t) \left\{ y_i(t) - \phi_i^T(t) \begin{bmatrix} \hat{\alpha}_{i-1}(t) \\ \hat{\theta}_i(t-1) \end{bmatrix} \right\}, \quad (137)$$

$$L_i(t) = \frac{P_{i-1}(t) \phi_i(t)}{1 + \phi_i^T(t) P_{i-1}(t-1) \phi_i(t)}, \quad (138)$$

$$\begin{aligned} P_i(t) &= [\mathbf{I} - L_i(t) \phi_i^T(t)] P_{i-1}(t), \\ \hat{\theta}_i(0) &= \mathbf{1}_{nr}/p_0, \quad i=2, 3, \dots, m, \end{aligned} \quad (139)$$

$$\phi(t) = [\mathbf{u}^T(t-1) \quad \mathbf{u}^T(t-2) \quad \dots \quad \mathbf{u}^T(t-n)]^T, \quad (140)$$

$$\psi(t) = [y(t-1) \quad y(t-2) \quad \dots \quad y(t-n)], \quad (141)$$

$$\psi(t) = \begin{bmatrix} \psi_1(t) \\ \psi_2(t) \\ \vdots \\ \psi_m(t) \end{bmatrix}, \quad \phi_i(t) = \begin{bmatrix} -\psi_i^T(t) \\ \phi(t) \end{bmatrix}. \quad (142)$$

PC-LS 算法的特点是: 各子系统协方差阵 $P_i(t) \in \mathbf{R}^{(n+nr) \times (n+nr)}$ 是耦合的. $\hat{\alpha}(t) := \hat{\alpha}_m(t)$, $\hat{\theta}_1(t)$, $\hat{\theta}_2(t)$, \dots , $\hat{\theta}_m(t)$ 分别是参数向量 α , θ_1 , θ_2 , \dots , θ_m 的估计.

PC-LS 算法实现步骤如下.

1) 初始化. $t=1$, $\hat{\alpha}_m(0) = \mathbf{1}_n/p_0$, $\hat{\theta}_1(0) = \mathbf{1}_{nr}/p_0$, $P_m(0) = p_0 \mathbf{I}_{n+nr}$, $i=1, 2, \dots, m$, $p_0 = 10^6$.

2) 收集输入输出数据 $u(t)$ 和 $y(t)$, 由式(140) 构成输入信息向量 $\varphi(t)$, 由式(141) 构成输出信息矩阵 $\psi(t)$, 从式(142) $\psi(t)$ 中提取 $\psi_i(t)$ ($i=1, 2, \dots, m$) 并构成 $\varphi_i(t)$.

3) 由式(135) 计算增益向量 $L_1(t)$, 由式(136) 计算协方差阵 $P_1(t)$, 由式(134) 刷新参数估计 $\hat{\alpha}_1(t)$ 和 $\hat{\theta}_1(t)$.

4) 对每一个 i ($i=2, 3, \dots, m$), 由式(138) 计算增益向量 $L_i(t)$, 由式(139) 计算协方差阵 $P_i(t)$, 由式(137) 刷新参数估计 $\hat{\alpha}_i(t)$ 和 $\hat{\theta}_i(t)$.

5) t 增 1 转到第 2 步.

注 6 PC-LS 算法(134) — (142) 不需要计算矩阵逆, 也不需要计算 $\hat{\alpha}_i(t)$ 的平均值, 而且 m 个协方

差阵 $P_i(t)$ 维数 $(n+nr) \times (n+nr)$ 要远小于多变量 RLS 算法 (109) — (111) 的协方差阵 $P(t)$ 的维数 $(n+mnr) \times (n+mnr)$. m 个协方差阵 $P_i(t)$ 包含了 $mn^2(r+1)^2$ 个元素, 远远小于协方差阵 $P(t)$ 的 $n^2(mr+1)^2$ 个元素. 当 $m=n=r=5$ 时, 前者为 $mn^2(r+1)^2=4\ 500$, 后者为 $n^2(mr+1)^2=16\ 900$.

5 有色噪声干扰的多变量系统耦合辨识方法

耦合辨识概念能够用于其他多变量系统辨识, 并且可以结合各种辨识方法, 如递推最小二乘方法、随机梯度辨识方法、辅助模型辨识方法、多新息辨识方法^[9]、递阶辨识方法(两阶段或多阶段辨识方法)^[5]、迭代辨识方法^[8, 39]、基于滤波的辨识方法^[22]、牛顿方法^[23]等, 研究各种干扰噪声模型的多变量系统辨识问题, 提出相应的耦合辨识方法.

对于有色噪声干扰的多变量系统, 可以针对系统模型结构和噪声模型的不同形式, 借用“增广”或“广义”化参数向量的维数的方法, 如增广随机梯度、增广最小二乘、广义随机梯度、广义最小二乘、广义增广随机梯度、广义增广最小二乘、辅助模型随机梯度或辅助模型最小二乘等^[7]方法, 来研究相应的耦合辨识方法. 这里的方法也可换为迭代辨识方法、多新息辨识方法等.

5.1 多变量伪线性回归系统的全耦合辨识方法

考虑下列有色噪声干扰的多变量伪线性回归系统

$$y(t) = \Phi(t) \theta + C^{-1}(z) D(z) v(t), \quad (143)$$

其中 $y(t) = [y_1(t) \ y_2(t) \ \dots \ y_m(t)]^T \in \mathbf{R}^m$ 为系统输出向量, $\Phi(t) \in \mathbf{R}^{m \times n}$ 是由系统输入输出数据构成的回归信息矩阵, $v(t) = [v_1(t) \ v_2(t) \ \dots \ v_m(t)]^T \in \mathbf{R}^m$ 是零均值白噪声向量, $\theta \in \mathbf{R}^n$ 是系统模型待辨识的参数向量, $C(z)$ 和 $D(z)$ 为单位阵, 或为单位后移算子 z^{-1} ($z^{-1}x(t) = x(t-1)$) 的多项式

$$c(z) := 1 + c_1 z^{-1} + c_2 z^{-2} + \dots + c_{n_c} z^{-n_c}, \quad c_i \in \mathbf{R},$$

$$d(z) := 1 + d_1 z^{-1} + d_2 z^{-2} + \dots + d_{n_d} z^{-n_d}, \quad d_i \in \mathbf{R}$$

或为多项式矩阵

$$C(z) = I + C_1 z^{-1} + C_2 z^{-2} + \dots + C_{n_c} z^{-n_c}, \quad C_i \in \mathbf{R}^{m \times m},$$

$$D(z) = I + D_1 z^{-1} + D_2 z^{-2} + \dots + D_{n_d} z^{-n_d}, \quad D_i \in \mathbf{R}^{m \times m}.$$

c_i 和 d_i 为噪声模型参数, $C_i \in \mathbf{R}^{m \times m}$ 和 $D_i \in \mathbf{R}^{m \times m}$ 为待辨识噪声模型的参数矩阵. 假设 n_c 和 n_d 已知, $t \leq 0$ 时 $y(t) = 0$, $\Phi(t) = 0$ 和 $v(t) = 0$.

也可在式 (143) 输出向量 $y(t)$ 前引入多项式 $\alpha(z)$ 或矩阵多项式 $A(z)$, 推广为一般情形

$$\alpha(z) y(t) = \Phi(t) \theta + C^{-1}(z) D(z) v(t), \quad (144)$$

或

$$A(z) y(t) = \Phi(t) \theta + C^{-1}(z) D(z) v(t), \quad (145)$$

其中 $\alpha(z)$ 的定义同上.

$$A(z) := I + A_1 z^{-1} + A_2 z^{-2} + \dots + A_{n_a} z^{-n_a}, \quad A_i \in \mathbf{R}^{m \times m}.$$

进一步可推广为

$$y(t) = \frac{\Phi(t)}{\alpha(z)} \theta + C^{-1}(z) D(z) v(t), \quad (146)$$

或

$$y(t) = A^{-1}(z) \Phi(t) \theta + C^{-1}(z) D(z) v(t). \quad (147)$$

同样 $C(z)$ 和 $D(z)$ 可为单位阵, 或多项式, 或多项式矩阵. 系统 (146) 和 (147) 需要借助于辅助模型辨识思想^[7]来研究其耦合辨识方法.

针对 $C(z)$ 和 $D(z)$ 的不同取值, 可以从式 (143) 得到下列几种特殊形式.

1) 多变量滑动平均伪线性系统 (multivariable moving average pseudo-linear system):

$$y(t) = \Phi(t) \theta + d(z) v(t), \quad (148)$$

或

$$y(t) = \Phi(t) \theta + D(z) v(t). \quad (149)$$

读者可以利用耦合辨识概念, 研究系统 (148) 和 (149) 的全耦合增广随机梯度辨识算法与全耦合增广最小二乘辨识算法, 以及基于滤波的全耦合随机梯度辨识算法与全耦合最小二乘辨识算法、全耦合迭代辨识方法等.

2) 多变量自回归伪线性系统 (multivariable autoregressive pseudo-linear system):

$$y(t) = \Phi(t) \theta + \frac{1}{c(z)} v(t), \quad (150)$$

或

$$y(t) = \Phi(t) \theta + C^{-1}(z) v(t). \quad (151)$$

能够推导系统 (150) 和 (151) 的全耦合广义随机梯度辨识算法、全耦合广义最小二乘辨识算法、全耦合多新息辨识算法等, 以及基于滤波的全耦合辨识算法.

3) 多变量自回归滑动平均伪线性系统 (multivariable autoregressive moving average pseudo-linear system):

$$y(t) = \Phi(t) \theta + \frac{d(z)}{c(z)} v(t), \quad (152)$$

或

$$y(t) = \Phi(t) \theta + \frac{D(z)}{c(z)} v(t), \quad (153)$$

或

$$y(t) = \Phi(t) \theta + C^{-1}(z) d(z) v(t), \quad (154)$$

或

$$y(t) = \Phi(t) \theta + C^{-1}(z) D(z) v(t). \quad (155)$$

能够推导出系统(152) — (155) 的全耦合广义增广随机梯度辨识算法与全耦合广义增广最小二乘辨识算法, 以及基于滤波的全耦合随机梯度辨识算法与全耦合最小二乘辨识算法等.

5.2 类多变量方程误差类系统的部分耦合辨识方法

考虑下列类多变量 CARARMA 系统(多变量 CARARMA-like 系统), 又称类多变量 ARARMAX 系统(多变量 ARARMAX-like 系统)^[5]

$$\alpha(z) y(t) = Q(z) u(t) + C^{-1}(z) D(z) v(t), \quad (156)$$

有关变量定义同上.

针对 $C(z)$ 和 $D(z)$ 的不同取值, 可以得到下列几种特殊形式.

1) 类多变量受控滑动平均系统(类多变量 CARMA 系统, 简称多变量 CARMA-like 系统, 或类多变量 ARMAX 系统, 简称多变量 ARMAX-like 系统. “类”翻译成“-like”其他类同):

$$\alpha(z) y(t) = Q(z) u(t) + d(z) v(t) \quad (157)$$

或

$$\alpha(z) y(t) = Q(z) u(t) + D(z) v(t). \quad (158)$$

读者可以利用耦合辨识概念, 研究系统(157) 和(158) 的部分耦合增广随机梯度辨识算法与部分耦合增广最小二乘辨识算法, 以及基于滤波的部分耦合随机梯度辨识算法与部分耦合最小二乘辨识算法等.

2) 类多变量受控自回归系统(多变量 CARAR-like 系统或多变量 ARARX-like 系统):

$$\alpha(z) y(t) = Q(z) u(t) + \frac{1}{c(z)} v(t) \quad (159)$$

或

$$\alpha(z) y(t) = Q(z) u(t) + C^{-1}(z) v(t). \quad (160)$$

能够推导出系统(159) 和(160) 的部分耦合广义随机梯度辨识算法与部分耦合广义最小二乘辨识算法, 以及基于滤波的部分耦合随机梯度辨识算法与部分耦合最小二乘辨识算法、迭代辨识方法等.

3) 类多变量受控自回归自回归滑动平均系统(多变量 CARARMA-like 系统或多变量 ARARMAX-like 系统):

$$\alpha(z) y(t) = Q(z) u(t) + \frac{d(z)}{c(z)} v(t), \quad (161)$$

或

$$\alpha(z) y(t) = Q(z) u(t) + \frac{D(z)}{c(z)} v(t), \quad (162)$$

或

$$\alpha(z) y(t) = Q(z) u(t) + C^{-1}(z) d(z) v(t), \quad (163)$$

或

$$\alpha(z) y(t) = Q(z) u(t) + C^{-1}(z) D(z) v(t). \quad (164)$$

能够推导出系统(161) — (164) 的部分耦合广义增广随机梯度辨识算法与部分耦合广义增广最小二乘辨识算法, 以及基于滤波的部分耦合随机梯度辨识算法与部分耦合最小二乘辨识算法等.

5.3 类多变量输出误差类系统的部分耦合辨识方法

下面讨论类多变量输出误差类系统的耦合辨识算法. 考虑下列类多变量 CARARMA 系统(多变量 CARARMA-like 系统), 又称类多变量 ARARMAX 系统(多变量 ARARMAX-like 系统)

$$y(t) = \frac{Q(z)}{\alpha(z)} u(t) + C^{-1}(z) D(z) v(t), \quad (165)$$

有关变量定义同上. 这类系统需要借助于辅助模型辨识思想^[7]来研究其耦合辨识方法.

针对 $C(z)$ 和 $D(z)$ 的不同取值, 从系统(165) 可以得到下列几种特殊形式.

1) 类多变量输出误差系统(multivariable OE-like system):

$$y(t) = \frac{Q(z)}{\alpha(z)} u(t) + v(t). \quad (166)$$

读者可以利用耦合辨识概念和辅助模型辨识思想, 研究系统(166) 的部分耦合辅助模型随机梯度辨识算法与部分耦合辅助模型最小二乘辨识算法.

2) 类多变量输出误差滑动平均系统(multivariable OEMA-like system):

$$y(t) = \frac{Q(z)}{\alpha(z)} u(t) + d(z) v(t) \quad (167)$$

或

$$y(t) = \frac{Q(z)}{\alpha(z)} u(t) + D(z) v(t). \quad (168)$$

读者可以推导出系统(167) 和(168) 的部分耦合辅助模型增广随机梯度辨识算法与部分耦合辅助模型增广最小二乘辨识算法, 以及基于滤波的部分耦合辅助模型随机梯度辨识算法与部分耦合辅助最小二乘辨识算法.

3) 类多变量输出误差自回归系统(multivariable OEAR-like system):

$$y(t) = \frac{Q(z)}{\alpha(z)} u(t) + \frac{1}{c(z)} v(t) \quad (169)$$

或

$$y(t) = \frac{Q(z)}{\alpha(z)}u(t) + C^{-1}(z)v(t). \quad (170)$$

能够推导出系统(169)和(170)的部分耦合辅助模型广义随机梯度辨识算法与部分耦合辅助广义最小二乘辨识算法,以及基于滤波的部分耦合辅助模型随机梯度辨识算法与部分耦合辅助最小二乘辨识算法.

4) 类多变量输出误差自回归滑动平均系统 (multivariable OEARMA-like system, i. e., multivariable Box-Jenkins-like system):

$$y(t) = \frac{Q(z)}{\alpha(z)}u(t) + \frac{d(z)}{c(z)}v(t), \quad (171)$$

或

$$y(t) = \frac{Q(z)}{\alpha(z)}u(t) + \frac{D(z)}{c(z)}v(t), \quad (172)$$

或

$$y(t) = \frac{Q(z)}{\alpha(z)}u(t) + C^{-1}(z)d(z)v(t), \quad (173)$$

或

$$y(t) = \frac{Q(z)}{\alpha(z)}u(t) + C^{-1}(z)D(z)v(t). \quad (174)$$

能够推导系统(171) — (174)的部分耦合辅助模型广义增广随机梯度辨识算法与部分耦合辅助模型广义增广最小二乘辨识算法,以及基于滤波的部分耦合辅助模型随机梯度辨识算法与部分耦合辅助模型最小二乘辨识算法.

5.4 多变量方程误差类系统的耦合辨识方法

下面简单给出可用耦合辨识概念来研究的多变量方程误差类系统模型结构.

考虑下列类多变量 CARARMA 系统(多变量 CARARMA-like 系统), 又称类多变量 ARARMAX 系统(多变量 ARARMAX-like 系统)

$$A(z)y(t) = B(z)u(t) + C^{-1}(z)D(z)v(t), \quad (175)$$

有关变量定义同上, 多项式矩阵 $B(z)$ 定义为

$$B(z) = B_1z^{-1} + B_2z^{-2} + \dots + B_{n_b}z^{-n_b}, \quad B_i \in \mathbf{R}^{m \times r}.$$

针对 $C(z)$ 和 $D(z)$ 的不同取值, 可以得到下列几种特殊形式.

1) 多变量方程误差系统(multivariable equation error system) 或多变量 CAR 系统(多变量 ARX 系统)

$$A(z)y(t) = B(z)u(t) + v(t). \quad (176)$$

2) 多变量方程误差滑动平均系统(multivariable equation error moving average system, i. e., multivariable EEMA system) 或多变量 CARMA 系统(多变量

ARMAX 系统):

$$A(z)y(t) = B(z)u(t) + d(z)v(t) \quad (177)$$

或

$$A(z)y(t) = B(z)u(t) + D(z)v(t). \quad (178)$$

3) 多变量方程误差自回归系统(multivariable equation error autoregressive system, i. e., multivariable EEAR system), 或多变量 CARAR 系统(多变量 ARARX 系统):

$$A(z)y(t) = B(z)u(t) + \frac{1}{c(z)}v(t) \quad (179)$$

或

$$A(z)y(t) = B(z)u(t) + C^{-1}(z)v(t). \quad (180)$$

4) 多变量受控自回归自回归滑动平均系统(multivariable CARARMA system):

$$A(z)y(t) = B(z)u(t) + \frac{d(z)}{c(z)}v(t), \quad (181)$$

或

$$A(z)y(t) = B(z)u(t) + \frac{D(z)}{c(z)}v(t), \quad (182)$$

或

$$A(z)y(t) = B(z)u(t) + C^{-1}(z)d(z)v(t), \quad (183)$$

或

$$A(z)y(t) = B(z)u(t) + C^{-1}(z)D(z)v(t). \quad (184)$$

5.5 多变量输出误差类系统的耦合辨识方法

下面简单给出可用耦合辨识概念来研究的多变量输出误差类系统模型结构.

考虑下列多变量 OEARMA 系统(多变量 Box-Jenkins 系统)

$$y(t) = A^{-1}(z)B(z)u(t) + C^{-1}(z)D(z)v(t). \quad (185)$$

进一步可广义化为

$$\zeta(z)y(t) = A^{-1}(z)B(z)u(t) + C^{-1}(z)D(z)v(t) \quad (186)$$

或

$$F(z)y(t) = A^{-1}(z)B(z)u(t) + C^{-1}(z)D(z)v(t), \quad (187)$$

其中

$$\zeta(z) = 1 + \zeta_1z^{-1} + \zeta_2z^{-2} + \dots + \zeta_nz^{-n}, \quad \zeta_i \in \mathbf{R},$$

$$F(z) = I + F_1z^{-1} + F_2z^{-2} + \dots + F_{n_f}z^{-n_f}, \quad F_i \in \mathbf{R}^{m \times m}.$$

针对 $C(z)$ 和 $D(z)$ 的不同取值, 从式(185)可以得到下列几种特殊形式.

1) 多变量输出误差系统(multivariable output error system):

$$y(t) = A^{-1}(z)B(z)u(t) + v(t). \quad (188)$$

读者可以利用耦合辨识概念, 研究系统(188)的部分耦合辅助模型随机梯度辨识算法与部分耦合辅助模型最小二乘辨识算法.

2) 多变量输出误差滑动平均系统(multivariable OEMA system) :

$$y(t) = A^{-1}(z) B(z) u(t) + d(z) v(t) \quad (189)$$

或

$$y(t) = A^{-1}(z) B(z) u(t) + D(z) v(t). \quad (190)$$

3) 多变量输出误差自回归系统(multivariable OEAR system) :

$$y(t) = A^{-1}(z) B(z) u(t) + \frac{1}{c(z)} v(t) \quad (191)$$

或

$$y(t) = A^{-1}(z) B(z) u(t) + C^{-1}(z) v(t). \quad (192)$$

4) 多变量输出误差自回滑动平均系统(multivariable OEARMA system) ,即多变量 Box-Jenkins 系统:

$$y(t) = A^{-1}(z) B(z) u(t) + \frac{d(z)}{c(z)} v(t) , \quad (193)$$

或

$$y(t) = A^{-1}(z) B(z) u(t) + \frac{D(z)}{c(z)} v(t) , \quad (194)$$

或

$$y(t) = A^{-1}(z) B(z) u(t) + C^{-1}(z) d(z) v(t) , \quad (195)$$

或

$$y(t) = A^{-1}(z) B(z) u(t) + C^{-1}(z) D(z) v(t). \quad (196)$$

6 结语

耦合辨识是基于多变量子系统间存在耦合参数而发展起来的一种新型辨识方法. 耦合辨识概念是解决结构复杂、参数耦合线性或非线性多变量系统辨识问题而提出的一种辨识理念与方法. 目的是通过子系统参数估计的耦合, 减少子系统参数向量的冗余估计, 减小辨识算法的计算量、提高辨识算法的收敛速度.

耦合辨识概念也可以用于研究一般矩阵方程和耦合矩阵方程的梯度迭代求解方法和最小二乘迭代求解方法. 本文讨论了多变量线性回归系统的几种全耦合最小二乘辨识算法、耦合随机梯度辨识算法, 讨论了拥有公分母特征多项式多变量系统的几种部分耦合最小二乘辨识方法、部分耦合随机梯度辨识算法. 最后, 列出了一些可用耦合辨识概念进行研究的有色噪声干扰的一些多变量系统模型结构, 包括有色噪声干扰多变量伪线性回归系统、多变量方程误差系统和多变量输出误差系统、类多变量方程误差系统和类多变量输出误差系统等.

提出的耦合辨识概念, 能够研究和解决国际著

名专家、加拿大的 Sinha 和 Kwong 教授在《Automatica》(1979年, 第4卷, 第471-475页) 中提出的存在交互干扰噪声的耦合多变量系统领域30多年的辨识难题.

参考文献

References

- [1] 丁锋. 鞅超收敛定理与遗忘因子最小二乘算法的收敛性分析[J]. 控制理论与应用, 1997, 14(1): 90-95
DING Feng. Martingale hyperconvergence theorem and convergence of the forgetting factor least squares algorithm [J]. Control Theory and Applications, 1997, 14(1): 90-95
- [2] 丁锋 杨家本. 关于鞅超收敛定理与遗忘因子最小二乘算法的收敛性分析[J]. 控制理论与应用, 1999, 16(4): 569-572
DING Feng, YANG Jiaben. Remarks on the martingale hyperconvergence theorem and convergence analysis of the forgetting factor least squares algorithms [J]. Control Theory and Applications, 1999, 16(4): 569-572
- [3] 丁锋 杨慧中 纪志成. 时变系统辨识方法及其收敛定理[J]. 江南大学学报: 自然科学版, 2006, 5(1): 115-126
DING Feng, YANG Huizhong, JI Zhicheng. Time-varying system identification methods and convergence theorems [J]. Journal of Southern Yangtze University: Natural Science Edition, 2006, 5(1): 115-126
- [4] 丁锋. 系统辨识(1): 辨识导引[J]. 南京信息工程大学学报: 自然科学版, 2011, 3(1): 1-22
DING Feng. System identification. Part A: Introduction to the identification [J]. Journal of Nanjing University of Information Science & Technology: Natural Science Edition, 2011, 3(1): 1-22
- [5] 丁锋. 系统辨识(2): 系统描述的基本模型[J]. 南京信息工程大学学报: 自然科学版, 2011, 3(2): 97-117
DING Feng. System identification. Part B: Basic models for system description [J]. Journal of Nanjing University of Information Science & Technology: Natural Science Edition, 2011, 3(2): 97-117
- [6] 丁锋. 系统辨识(3): 辨识精度与辨识基本问题[J]. 南京信息工程大学学报: 自然科学版, 2011, 3(3): 193-226
DING Feng. System identification. Part C: Identification accuracy and basic problems [J]. Journal of Nanjing University of Information Science & Technology: Natural Science Edition, 2011, 3(3): 193-226
- [7] 丁锋. 系统辨识(4): 辅助模型辨识思想与方法[J]. 南京信息工程大学学报: 自然科学版, 2011, 3(4): 289-318
DING Feng. System identification. Part D: Auxiliary model identification idea and methods [J]. Journal of Nanjing University of Information Science & Technology: Natural Science Edition, 2011, 3(4): 289-318
- [8] 丁锋. 系统辨识(5): 迭代搜索原理与辨识方法[J]. 南京信息工程大学学报: 自然科学版, 2011, 3(6):

- 481-510
DING Feng. System identification. Part E: Iterative search principle and identification methods [J]. Journal of Nanjing University of Information Science & Technology: Natural Science Edition 2011 3(6):481-510
- [9] 丁锋. 系统辨识(6):多新息辨识理论与方法[J]. 南京信息工程大学学报:自然科学版 2012 4(1):1-28
DING Feng. System identification. Part F: Multi-innovation identification theory and methods [J]. Journal of Nanjing University of Information Science & Technology: Natural Science Edition 2012 4(1):1-28
- [10] 丁锋. 系统辨识(7):递阶辨识原理与方法[J]. 南京信息工程大学学报:自然科学版 2012 4(2):97-124
DING Feng. System identification. Part G: Hierarchical identification principle and methods [J]. Journal of Nanjing University of Information Science & Technology: Natural Science Edition 2012 4(2):97-124
- [11] 丁锋. 基于输出估计的多输入系统随机梯度估计算法[J]. 南京信息工程大学学报:自然科学版,2010,2(6):481-488
DING Feng. Stochastic gradient estimation algorithm for multiple-input systems based on the output estimation [J]. Journal of Nanjing University of Information Science & Technology: Natural Science Edition,2010,2(6):481-488
- [12] 丁锋, 萧德云, 丁韬. 多新息随机梯度辨识方法[J]. 控制理论与应用 2003 20(6):870-874
DING Feng, XIAO Deyun, DING Tao. Multi-innovation stochastic gradient identification methods [J]. Control Theory and Application 2003 20(6):870-874
- [13] Ding F, Liu G, Liu X P. Partially coupled stochastic gradient identification methods for non-uniformly sampled systems [J]. IEEE Transactions on Automatic Control, 2010 55(8):1976-1981
- [14] Bristow D A, Tharayil M, Alleyne A G. A survey of iterative learning control [J]. IEEE Control Systems Magazine 2006 26(3):96-114.
- [15] Ding F, Chen T. Hierarchical gradient-based identification of multivariable discrete-time systems [J]. Automatica 2005 41(2):315-325
- [16] Ding F, Chen T. Hierarchical least squares identification methods for multivariable systems [J]. IEEE Transactions on Automatic Control 2005 50(3):397-402
- [17] Wang L Y, Ding F, Liu X P. Consistency of HLS estimation algorithms for MIMO ARX-like systems [J]. Applied Mathematics and Computation 2007 190(2):1081-1093
- [18] 丁锋, 萧德云. 多变量系统状态空间模型的递阶辨识[J]. 控制与决策 2005 20(8):848-853 859
DING Feng, XIAO Deyun. Hierarchical identification of state space models for multivariable systems [J]. Control and Decision 2005 20(8):848-853 859
- [19] Ding F, Chen T. Hierarchical identification of lifted state-space models for general dual-rate systems [J]. IEEE Transactions on Circuits and Systems. I: Regular Papers, 2005 52(6):1179-1187
- [20] Han H Q, Xie L, Ding F, et al. Hierarchical least squares based iterative identification for multivariable systems with moving average noises [J]. Mathematical and Computer Modelling 2010 51(9/10):1213-1220
- [21] Zhang Z N, Ding F, Liu X G. Hierarchical gradient based iterative parameter estimation algorithm for multivariable output error moving average systems [J]. Computers & Mathematics with Applications 2011 61(3):672-682
- [22] Wang D Q, Ding F. Input-output data filtering based recursive least squares parameter estimation for CARARMA systems [J]. Digital Signal Processing 2010 20(4):991-999
- [23] Ding F, Liu X P, Liu G. Identification methods for Hammerstein nonlinear systems [J]. Digital Signal Processing 2011 21(2):215-238
- [24] 方崇智, 萧德云. 过程辨识 [M]. 北京: 清华大学出版社, 1988
FANG Chongzhi, XIAO Deyun. Process Identification [M]. Beijing: Tsinghua University Press, 1988
- [25] 丁锋, 杨家本. 大系统的递阶辨识 [J]. 自动化学报, 1999 25(5):647-654
DING Feng, YANG Jiaben. Hierarchical identification of large scale systems [J]. Acta Automatica Sinica, 1999 25(5):647-654
- [26] Ding F, Yang J B, Xu Y M. Convergence of hierarchical stochastic gradient identification for transfer function matrix models [J]. Control Theory and Applications, 2001, 18(6):949-953
- [27] Ding F, Chen T. Performance analysis of multi-innovation gradient type identification methods [J]. Automatica, 2007 43(1):1-14
- [28] Ding F, Liu X P, Liu G. Multi-innovation least squares identification for system modeling [J]. IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics, 2010 40(3):767-778
- [29] Ding F. Several multi-innovation identification methods [J]. Digital Signal Processing 2010 20(4):1027-1039
- [30] Wang D Q, Ding F. Performance analysis of the auxiliary models based multi-innovation stochastic gradient estimation algorithm for output error systems [J]. Digital Signal Processing 2010 20(3):750-762
- [31] Liu Y J, Yu L, Ding F. Multi-innovation extended stochastic gradient algorithm and its performance analysis [J]. Circuits, Systems and Signal Processing, 2010, 29(4):649-667
- [32] Golub G H, Van Loan C F. Matrix Computations [M]. 3rd Ed. Baltimore, MD: Johns Hopkins Univ Press, 1996
- [33] Ding F, Liu Y J, Liu X G, et al. Coupled least squares identification for multivariable systems [J]. IET Control Theory (Submitted)
- [34] 丁锋, 丁韬, 杨家本, 等. 时变参数遗忘梯度估计算法的收敛性 [J]. 自动化学报 2002 28(6):962-968
DING Feng, DING Tao, YANG Jiaben, et al. Convergence of forgetting gradient estimation algorithm for time-varying parameters [J]. Acta Automatica Sinica, 2002, 28(6):962-968
- [35] Ding J, Shi Y, Wang H G, et al. A modified stochastic gradient based parameter estimation algorithm for dual-rate sampled-data systems [J]. Digital Signal Processing, 2010 20(4):1238-1247
- [36] Chen J, Ding F. Modified stochastic gradient algorithms

- with fast convergence rates [J]. *Journal of Vibration and Control* 2011, 17(9): 1281-1286
- [37] Liu Y J, Sheng J, Ding R F. Convergence of stochastic gradient algorithm for multivariable ARX-like systems [J]. *Computers & Mathematics with Applications* 2010, 59(8): 2615-2627
- [38] Liu Y J, Ding R. Partially coupled estimation algorithm for discrete-time multiple-input multiple-output systems [C]//The 2012 24th Chinese Control and Decision Conference (2012 CCDC), Taiyuan, China 2012: 2093-2098
- [39] Li J H, Ding R F, Yang Y. Iterative parameter identification methods for nonlinear functions [J]. *Applied Mathematical Modelling* 2012, 26(6): 2739-2750

System identification. Part H: Coupled identification concept and methods

DING Feng^{1, 2, 3}

1 School of Internet of Things Engineering, Jiangnan University, Wuxi 214122

2 Control Science and Engineering Research Center, Jiangnan University, Wuxi 214122

3 Key Laboratory of Advanced Process Control for Light Industry (Ministry of Education), Jiangnan University, Wuxi 214122

Abstract Coupled identification is an important branch of system identification and is a new identification concept which is used mainly to study identification problems of linear and nonlinear multivariable systems with complex structures and parameter coupling. The auxiliary model identification idea, the multi-innovation identification theory, the hierarchical identification principle and the coupled identification concept are new identification research ideas, concepts and methods and can be used to study identification problems of systems with unknown process variables, to improve the convergence rates and accuracies of identification methods, to solve identification problems of large-scale multivariable systems with complex structures and of multivariable systems with parameter coupling, reducing the computational load of the identification algorithms.

This paper introduces the coupled identification concept of multivariable systems, discusses the (full) coupled least squares identification methods, the (full) coupled stochastic gradient identification methods, the partially coupled stochastic gradient identification methods, the partially coupled least squares identification methods etc for multivariable systems. Finally, we show that the coupled identification methods can be applied to multivariable systems with colored noises, list some model structures of some multivariable systems and indicate that the coupled identification concept can combine the auxiliary model identification idea, the multi-innovation identification theory, the hierarchical identification principle, the iterative search principle (the gradient iteration, the least squares iteration, the Newton iteration) to study identification problems of linear and nonlinear multivariable systems with colored noises.

Key words iterative identification; recursive identification; parameter estimation; FIR model; equation error model; CAR model; CARMA model; CARAR model; CARARMA model; output error model; OEMA model; OEAR model; auxiliary model identification; multi-innovation identification; hierarchical identification; coupled identification